Causality in medicine with particular reference to the viral causation of cancers

Brendan Owen Clarke

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

UCL

Department of Science and Technology Studies

UCL

August 20, 2010

Revised version January 24, 2011

I, Brendan Owen Clarke, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

.....

January 24, 2011

Acknowledgements

This thesis would not have been written without the support and inspiration of Donald Gillies, who not only supervised me, but also introduced me to history and philosophy of science in the first instance. I have been very privileged to be the beneficiary of so much of his clear thinking over the last few years. Donald: thank you for everything, and I hope this thesis lives up to your expectations.

I am also extremely grateful to Michela Massimi, who has acted as my second supervisor. She has provided throughout remarkably insightful feedback on this project, no matter how distant it is from her own research interests. I'd also like to thank her for all her help and guidance on teaching matters.

I should also thank Vladimir Vonka for supplying many important pieces of both the cervical cancer history, as well as his own work on causality in medicine from the perspective of a real, working medical researcher.

Phyllis McKay-Illari provided a critical piece of the story, as well as many interesting and stimulating discussions about the more philosophical aspects of this thesis. Similarly, I also wish to thank the other members of the Causality reading group (Jonathan Everett, Toby Friend, David Lagnado, John Pemberton, Christoph Schulz and Adam White) for all the good-natured arguments that we have had.

The intellectual environment is incredibly important when producing such an interdisciplinary work. I'd therefore like to thank the students and staff of the department of Science and Technology Studies at UCL for their support and ideas. Particular thanks must go to Joe Cain, for reminding me about biology and teaching, as well as his critical eye on the first parts of the thesis, David Teira for all his acumen, Silvia De Bianchi for her help with Kant, and Chiara Ambrosio for reminding me that there's more to philosophy than philosophy of science. Here, I might also thank Hasok Chang for showing how beautiful integrated history and philosophy of science can be. Finally, I must acknowledge the virtues of Ian James Kidd for showing me how to make specialist research engaging.

These acknowledgements would not be complete without vigorously thanking Jon Williamson and Federica Russo for providing the philosophical germ of this thesis, as well as all their helpful comments on my earlier work. I wish to thank Carl Craver for suggesting that the relationship between evidence and mechanism was one worth exploring.

I'd also like to thank some individuals who significantly contributed to my medical education, in particular Kevin Brown and his colleagues at the Health Protection Agency, who taught me about the practical difficulties of relating clinical disease to newly discovered viruses, and Alan Bates, who made some important comments about post-mortem causality.

After all these expressions of scholarly gratitude, I cannot conclude without some more personal thanks. To Alex Durell, Neil Forrest, Nick Heath, Ross Howlett, Chris Matcham, Mark Richards, Martin Wallwork: thanks, chaps. Classificatorily fuzzy thanks must also go to Carl Pappenheim, not only for his help with causal graphs, but (much more importantly) everything else over the last two decades. Carl: if I thank you here, will you stop moaning about the wedding? Kelvin Meek should be remembered for helping with the more sophisticated historical aspects of this work, and for his friendship.

To my parents, my brother and my grandparents: thank you for absolutely everything over what has been, at times, a troublesome university career. To Peter and Lindie Allen, thank you for all your support.

Finally, this thesis is dedicated to my wife Rosemary and my son Thomas. As well as all the practical help with the preparation and production of the thesis itself, I could not have done any of this without you. Thank you.

London, August 18, 2010

Abstract

In this thesis, I give a metascientific account of causality in medicine. I begin with two historical cases of causal discovery. These are the discovery of the causation of Burkitt's lymphoma by the Epstein-Barr virus, and of the various viral causes suggested for cervical cancer. These historical cases then support a philosophical discussion of causality in medicine. This begins with an introduction to the Russo-Williamson thesis (RWT), and discussion of a range of counter-arguments against it. Despite these, I argue that the RWT is historically workable, given a small number of modifications. I then expand Russo and Williamson's account. I first develop their suggestion that causal relationships in medicine require some kind of evidence of mechanism. I begin with a number of accounts of mechanisms and produce a range of consensus features of them. I then develop this consensus position by reference to the two historical case studies with an eye to their operational competence. In particular, I suggest that it is mechanistic models and their representations which we are concerned with in medicine, rather than the mechanism as it exists in the world.

I then employ these mechanistic models to give an account of the sorts of evidence used in formulating and evaluating causal claims. Again, I use the two human viral oncogenesis cases to give this account. I characterise and distinguish evidence of mechanism from evidence of difference-making, and relate this to mechanistic models. I then suggest the relationship between types of evidence presents us with a means of tackling the reference-class problem.

This sets the scene for the final chapter. Here, I suggest the manner in which these two different classes of evidence become integrated is also reflected in the way that developing research programmes change as their associated causal claims develop.

Contents

Glossary and list of acronyms **12**

1 Introduction **13**

1.1 Non-human viral oncogenesis 15

1.2 Medical schemes of causality 17

1.3 A note on causal discovery 20

2 Epstein-Barr Virus and Burkitt's Lymphoma **25**

2.1 Introduction 25

2.2 Suspicion: BL as a tumour syndrome 28

2.3 Domain finding: BL as a geographical disease 33

2.4 Making mechanisms: EBV 42

2.5 Application: prospective epidemiology 51

3 Cervical Cancer **53**

3.1 Introduction 53

3.2 Suspicion: Cervical cancer as a venereal disease 56

3.3 Domain finding: HSV and cervical cancer 60

3.4 Making mechanisms I: HSV (1966-1983) 61

3.5 Application I: HSV 63

3.6 Making mechanisms II: HPV (1983–) 66

3.7 Application II: HPV 74

4 The Russo-Williamson Thesis **76**

4.1 Introduction 76

4.2 The Russo-Williamson Thesis 77

4.3 Philosophical counterarguments to the RWT 86

4.4 The RWT as an empirical proposition 92

4.5 Historical counterexamples to the RWT 95

5 Mechanisms **108**

5.1 Introduction 108

5.2 Mechanistic theories of causality . . . 109
5.3 Mechanisms in medicine . . . 126
5.4 Levels and gaps . . . 133
5.5 Functional definition . . . 140
5.6 Branching and track-switching . . . 142
5.7 Two kinds of effects . . . 143
5.8 Preventions and omissions . . . 146
5.9 A definition of mechanisms for the Russo-Williamson Thesis . . . 150

6 Evidence for Causality **151**
6.1 Introduction . . . 151
6.2 Two types of types of evidence . . . 152
6.3 Evidence in the cases . . . 154
6.4 Evidence of mechanism . . . 164
6.5 Evidence of difference-making . . . 170
6.6 Difference-making and the Reference Class Problem . . . 174

7 Conclusion: The Russo-Willamson Thesis Redux **179**
7.1 The Russo-Williamson Thesis redux . . . 179
7.2 Integration of evidence in practice . . . 181
7.3 Final conclusions . . . 182

Bibliography **183**

List of Figures

2.1 BL cases from the 1961 postal survey . . . 38
2.2 BL cases and climatic conditions . . . 39
2.3 BL cases and tsetse distribution . . . 40
2.4 BL cases and malaria distribution . . . 41
2.5 Route of the tumour safari . . . 41

3.1 HPV phylogenetic tree . . . 54

4.1 Simplified mechanism of glucose-glycogen interconversion . . . 96
4.2 Plausible causes of HCC . . . 103

5.1 Woodward's intervention condition M1 . . . 118
5.2 Woodward's intervention condition M2 . . . 119
5.3 Woodward's intervention condition M3 . . . 119
5.4 Woodward's intervention condition M4 . . . 119
5.5 Simple physiological mechanism . . . 143
5.6 Physiological mechanistic model . . . 144
5.7 Partial physio-pathological mechanistic model . . . 144
5.8 Physio-pathological mechanistic model . . . 145
5.9 Causes of divergence at *A* . . . 145
5.10 Preventing *D'* by preventing *X* and *Y* by intervention *I* . . . 146
5.11 Negative causation example 1 . . . 147
5.12 Negative causation example 2 . . . 147
5.13 Negative causation example 3 . . . 148

6.1 Cervical cancer example . . . 177

List of Tables

2.1 Summary of early BL cell lines 46

3.1 Summary of early types of HPV 72

Glossary

arbovirus a type of virus transmitted by insects. 47

leukaemia blood-cell tumours which exist in a distributed form in the blood and bone marrow. 17

lymphoma a solid tumour of some kinds of white blood-cells. 15, 17

metascientific deriving from the study of science or scientific practice. 14

reductionistic theories of causality that necessarily seek to reduce macro-level causal phenomena to the operations of micro-level interactions. 109

reductivist theories of causality that seek to define causal concepts in terms of non-causal ones. 108

sarcoma a tumour derived from connective tissue. 16

viral oncogenesis the process by which certain viruses can cause tumours. 14, 15

Acronyms

BL Burkitt's lymphoma. 17, 21, 25–29, 31, 32, 34–51, 53, 55, 61–64, 76, 78, 93–95, 130, 157–163, 165, 167, 169–172, 174, 175, 181, 182

BPV bovine papillomavirus. 16, 66, 68, 71

CFF cell-free filtrates. 15, 16, 66

CMI cell-mediated immunity. 103, 104

CMV cytomegalovirus. 61, 94

CNID chronic non-infectious diseases. 18, 19

CPE cytopathic effects. 17, 60, 64, 65, 93

CQ conserved quantities. 82, 110, 135, 137

CRIGs change-relating invariant generalisations. 121, 122, 124

CRPV cottontail rabbit papillomavirus. 16, 66, 68

DKG diathermoelectro-coagulation of the cervix. 59, 173, 177, 178

DME difference-making evidence. 153, 173

DSM Diagnostic and Statistical Manual. 156

EBM Evidence-Based Medicine. 84–86, 88

EBV Epstein-Barr Virus. 16, 21, 25–28, 42, 45–51, 53, 61–64, 76–78, 93–95, 151, 154, 157, 158, 161–166, 169–174, 180–182

EDM evidence of difference-making. 153, 170–174, 178–180, 182

EM evidence of mechanism. 153, 164, 165, 172, 173, 182

EV epidermodysplasia verruciformis. 54, 66–68, 70–72

G-1-P glucose-1-phosphate. 96

G-6-P glucose-6-phosphate. 96

GPD glyceraldehyde phosphate dehydrogenase. 97

HBV hepatitis B virus. 63, 76, 101–105, 129, 148, 149

HCC hepatocellular carcinoma. 63, 100–105, 148, 149, 161, 180

HKP Henle-Koch postulates. 13, 18, 19, 93

HPV human papillomavirus. 54, 55, 65–75, 77, 78, 138, 139, 141, 151, 154, 164, 165, 168, 182

HSV herpes simplex virus. 15, 42, 44, 47–49, 53, 54, 60–66, 69, 77, 82, 86, 93, 94, 105–107, 161, 163, 166, 167, 169–173, 177, 182

IM infectious mononucleosis. 27, 28, 50

MDC Machamer, Darden, Craver. 110–114, 116, 122, 124, 126, 127, 129, 132, 150, 158

ME mechanistic evidence. 153, 164–166, 172, 173

MMTV murine mammary tumour virus. 16

MT mark-transmission. 110, 133–135, 137

NMP New Mechanistic Philosophy. 110, 113, 123

p53 tumour protein 53. 55, 73

PCR polymerase chain reaction. 74, 100

PGM phosphoglycomutase. 96

pRb retinoblastoma gene product. 55, 73

RaHV-1 ranid herpesvirus 1. 17, 129

RCP reference class problem. 151, 174, 175, 177, 178

RCT randomised control trial. 84–86, 88, 89, 91

RWT Russo-Williamson thesis. 76, 77, 85–93, 97–101, 104, 105, 108, 109, 126–128, 142, 150–152, 154, 158, 160, 168, 175, 178, 179, 182, 183

SV40 Simian vacuolating virus 40. 17

UDPG uridine diphosphate glucose. 96

VCA EBV viral capsid antigen. 94, 95

VLP virus-like particles. 16, 17, 43–48, 66, 75, 102

Chapter 1

Introduction

In this thesis, I wish to explore the concept of causality in medicine. Rather than giving an account that focuses on theoretical aspects of this issue, I wish to concentrate on the operational features of causes in medical practice. How can we practically differentiate causal from non-causal processes? What is the relationship between evidence and causality? How can we best exploit causal knowledge to guide medical practice?

Why focus on these issues? First, causal knowledge brings advantages to medical practice that non-causal knowledge does not. These advantages are apparent in core medical practices, such as diagnosis and treatment, in medical research, such as disease and treatment discovery, and in the wider medical milieu, such as medicolegal and policy applications. Knowledge of causes supports these various practices because of the range of special properties that only causes possess. Causal relationships tend to be stable, or at least robust, over a range of background conditions. Even if they are not, it tends to be the case that we can predict which background conditions they are likely to remain stable over in a way that we cannot reliably do for non-causal relationships. This means that causal knowledge of, say, the aetiology of a disease makes for sounder judgements in a range of situations over non-causal knowledge. Causes are also exploitable: that is, they allow us to plan and perform interventions in a rational manner. In other words, good medical practice requires the notion of causality.

However, despite the great advantages of thinking causally, it has proven very difficult to construct practical schemes for determining cause that are suitable for general use in medicine. Physicians and medical researchers have made a great many attempts to devise such schemes, as I review in section 1.2. These can be, like the Henle-Koch postulates (HKP),[1] locally very successful. But it is uncontroversial to say that no one of these schemes provides a general way of detecting a causal relationship that has universal (or even broad) applicability. These practical difficulties are matched by similarly daunting theoretical ones. Philosophers too have produced a great number of theories of causality. Yet here too there is no dominant view or consensus. Nor is there a theory, or type of theory, that seems to satisfy even a modest desideratum of theoretical causal knowledge. So it is not the case that the practical problems that medical researchers face in thinking about causes arise from a lack of philosophical knowledge.

[1] I have provided a list of acronyms used in this thesis above. As with items in the bibliography, each is followed by the page number(s) upon which it appears.

However, because of the advantages of providing practically workable schemes of cause, the provision of causal tools suitable for clinical use seems, to me, a more pressing problem than the theoretical one. This means that I will give an account of causality that concentrates on how we should work with, and think about causes, rather than one that attempts to understand causality itself. In other words, in what follows I will take causality as a primitive concept that we might employ in order to do things in practice. These things include being able to discriminate causal from non-causal processes in a reliable way, to provide tools to understand and evaluate causal claims in practice and to take account of the relationship between evidence and cause. Importantly, I want to do this in a way that is generalisable, and does not depend upon, say, interpreting causes using any particular medical technology.[2] In short, I am interested in theories of causality that show how a particular causal phenomenon is situated within the causal structure of the world (Craver, 2007, 21), rather than one that attempts to define just what causality *is*.

Because this thesis is motivated by practical concerns, I will take an approach that might be broadly characterised as *metascientific*. Instead of beginning with a metaphysical stance regarding causality, and attempting to translate this position such that it is compatible with scientific practice, I will begin with the study of scientific practice itself. This study will be an historical one. In chapters 2 and 3 I will introduce two detailed historical examples in which complex causal claims were formulated and tested. Both of these cases are early examples of tumours caused by viral infections (viral oncogenesis). As such, they not only deal with the construction and evaluation of novel, specific, causal claims linking a particular virus with a particular clinical disease, but also demonstrate the much broader shift in causal thinking that, in part, arose from the discovery of these sorts of causes in medicine. Here, I will demonstrate the ways in which different types of evidence were generated and synthesised to support these new kinds of causal claim.

These case-studies will then form the foundations for the more philosophical parts of the thesis. I will begin this by exploring Russo and Williamson's recent work on causality in medicine. I argue that their distinction between causality and the types of evidence that support causal claims is correct. I then attempt to expand upon their distinction between mechanistic and probabilistic evidence. This programme of expanding upon their thesis, by means of recent historical episodes in medicine, forms the latter parts of the thesis.

For now, though, I need to introduce three key topics that will form the background to what follows. Therefore, in section 1.1, I will give some very short historical details about the discovery of viral oncogenesis in non-human animals. In section 1.2 I will then give a broad overview of the ways that medical researchers have attempted to understand causality. Finally, in section 1.3, I will give a short introduction to an account of causal discovery that, I think, is compatible with what follows.

[2] See footnote 8 on p.18 for an example of how schemes of causality that depend on particular technologies can be practically problematic.

1.1 Non-human viral oncogenesis

Cancer was historically held to be a transmissible disease. Yet with the introduction of the germ theory of disease in the late nineteenth century, an infectious origin for it became implausible (Epstein, 2001a, 413-4). However, in the middle decades of the twentieth century, the study of a number of tumours in animals gradually led to a re-evaluation of this concept. In time, several instances of viral oncogenesis - tumours caused by viruses - were discovered in non-human animals. These conditions were very important to researchers studying later cases of human viral oncogenesis for two reasons. First, they acted as exemplars of how viruses might generally act to cause cancer. Thus, the pattern of viral causation of cancers, and its discovery in these cases, was important as a methodological guide in the cases of human viral oncogenesis discussed in chapters 2 and 3. Second, several of these conditions played a more direct evidential role in causal discovery. For example, the knowledge that several different herpesvirus infections were capable of causing cancers in animals played a vital, but problematic, role in supporting the claim that infection with herpes simplex virus (HSV) caused cervical cancer. In this section, I will give a brief, non-technical, outline of these early examples of non-human viral oncogenesis.[3]

One general aspect of these cases is worth highlighting in advance. While a number of conditions, later determined to be virally caused, were clinically described in the early decades of the twentieth century, the actual viruses responsible were not discovered until much later. In fact, most of the viruses involved in the following examples were not described until after about 1950. This lag appears to be due to the technology required to give convincing descriptions of viruses. For example, the electron microscope, which played a vital role in discovering and describing oncoviruses, did not enter general use until the early 1950s.[4] What suspicions there were before this date were supported with a range of techniques that presented much less convincing evidence for the nature and characteristics of oncoviruses than the striking images produced by electron microscopy. For instance, a common technique was the transmission of tumours by cell-free filtrates (CFF).[5] This delay presents the historian with difficulties in precisely dating the beginnings of theories of viral oncogenesis. With that proviso in mind, I will review the important examples of non-human viral oncogenesis that informed the early work on human viral oncogenesis.

Marek disease. This is a T-cell lymphoma that affects chickens (Biggs, 2001). The disease was first clinically described by Marek in 1907 (Marek, 1907), and was recognised to be a malignant condition by Pappenheimer et al. in 1929 (Pappenheimer et al., 1929a). It was discovered that the disease could be transmitted by the inoculation of tissue from affected individuals into unaffected ones (Pappenheimer

[3] There are a number of good technical reviews of this material. See, for example, zur Hausen, 1980; zur Hausen, 2006, chapter 1 and Javier and Butel, 2008.

[4] Although Ernst Ruska and Max Knoll had developed an experimental electron microscope in 1931, it was not until 1939 that a transmission electron microscope (the '*Siemens Super*') was commercially available. Even by the beginning of 1945 only about 35 laboratories in the world were equipped with one (Ruska, 1986), and it was not until 1940 that reviews of viral detection were published (Kruger et al., 2000, 1714). This technology did not become widely disseminated outside Germany until the end of the Second World War, and it took some time before electron microscopy was a routine part of virological practice (Bang, 1952).

[5] This involved attempting to transmit the tumour by inoculations of tumour material that had been passed through a filter that was sufficiently fine to exclude any whole cells.

et al., 1929b). The causal virus (now known to be gallid herpesvirus 2) was discovered in 1967 (Churchill and Biggs, 1967) and described in 1969 (Witter et al., 1969). Interestingly, Epstein, who would later play such an important role in discovering Epstein-Barr Virus (EBV), was involved in early work on the description of this herpesvirus (Epstein et al., 1968).

Rous sarcoma. This disease, which causes sarcomas in chickens, was described and discovered to be transmissible by Rous in 1911 (Rous, 1911a; Rous, 1911b). A possible oncogenic mechanism, implicating an RNA virus in the development of this condition, was demonstrated in 1968 (Temin, 1968). This agent was later discovered to be a retrovirus by the discovery of virally-encoded reverse transcriptase activity. This was independently discovered by both Temin and David Baltimore in 1970 (Baltimore, 1970; Temin and Mizutani, 1970), leading to them jointly winning the Nobel Prize in Medicine in 1975 (Anonymous, 1975).

The causal virus is now known to be a member of the avian sarcoma-leukosis virus group. In addition to the Rous sarcoma virus, this group includes a further chicken virus which played a minor role in the early story of non-human viral oncogenesis. This is the **avian leukosis virus**, which was the agent responsible for one of the first known transmissible tumours (Ellermann and Bang, 1909). The complexities of this story are well reviewed in Pitot, 1983.

Shope papillomavirus. This papillomatous disease, affecting rabbits, was clinically described and shown to be transmissible by CFF in the early 1930s (Shope, 1932; Shope and Hurst, 1933). It was later demonstrated to be a maligant disease (Rous and Beard, 1934). The causal agent is now known to be a papillomavirus, cottontail rabbit papillomavirus (CRPV) (Ito and Evans, 1961). I discuss this disease in more detail in section 3.6.1.3, along with another papillomavirus infection of historical importance, **bovine papillomavirus (BPV)**.

Murine mammary tumour virus (MMTV). Mammary tumours in mice were discovered to be transmissible via milk (Bittner, 1936) and this transmission was suggested to be due to the action of a virus (Andervont and Bryan, 1944). Virus-like particles (VLP) were subsequently discovered in mouse milk using electron microscopy (Passey et al., 1950) and were noted to be similar to those seen in mammary tumour tissue biopsies. This agent is now known to be a milk-transmitted retrovirus called MMTV, and was, for a time, suggested as a possible cause of human disease (Gross et al., 1952). Perhaps because of this, murine models of breast cancers have played a vital role in understanding human breast cancers (Callahan and Smith, 2008, which prefaces an entire journal issue dedicated to this issue).

Lucké renal adenocarcinoma. This is a kidney tumour that affects the northern Leopard frog (*Rana pipiens*). It was described, noted to be transmissible by grafting and suspected to be of viral origin by Baldwin Lucké (Lucké, 1934; Lucké, 1938a; Lucké, 1938b). The suspicion of its viral origin was largely based on the presence of intranuclear inclusion bodies in tumour cells (see, e.g. Lucké, 1938a, 457). It was also noted that tumour cells could be maintained in culture, both *in vitro* (Lucké, 1939) and *in vivo* (Schlumberger and Lucké, 1949). Lucké continued to argue persuasively, on rather circumstantial evidence, that the tumour was caused by a virus (Lucké, 1952). However, it was not until the demonstration of herpesvirus-like particles in tumour cells (Fawcett, 1956) that any stronger

evidence became available. However, it appeared that a variety of different viral agents were present in these tumours (Gravell, 1971), and it took a number of years until the causal role of ranid herpesvirus 1 (RaHV-1) in the development of the tumour was clear (Naegele and Granoff, 1977).

Murine leukaemia and lymphoma. VLPs were demonstrated in certain types of mouse leukaemia during the 1950s and 1960s (Gross, 1952; Rauscher, 1962). A range of possible viral agents were suspected (comprehensively reviewed in Moloney, 1962). Similarly, a large number of viral agents were also detected in various murine lymphomas. Of interest here was the apparent role played by reovirus 3, and its relationship to Burkitt's lymphoma (Papadimitriou, 1966; Stanley et al., 1966; Stanley and Keast, 1967). I go on to discuss this issue in section 2.4.2.4.

Simian vacuolating virus 40 (SV40). Preparations of renal cells from Rhesus and cynomolgus monkeys yielded a polyomavirus (Sweet and Hilleman, 1960). This had no cytopathic effects (CPE) in its host species, but was found to induce lysis in cells derived from other primates. It was found to induce tumours when inoculated into hamsters (Eddy, 1962). Because rhesus monkey cells were used in the production of the Sabin polio vaccine, many millions of people were inadvertently exposed to SV40 during the second half of the twentieth century. This is worrying, as SV40 is thought to play a causal role in several human tumours (Martini et al., 2007).

Adenovirus. Adenoviruses usually cause fairly mild eye, respiratory tract and gut infections in humans. However, they were discovered to have oncogenic properties when transfected into non-human animals. Adenovirus type 12 was intially found to cause tumours in hamsters (Trentin et al., 1962). Similar properties were also described for adenovirus types 18 (Huebner et al., 1962), 7 (Girardi et al., 1964) and 31 (Pereira et al., 1965).

Herpes ateles* and *Herpes samirii These two primate herpesviruses were discovered to cause lymphomas when transfected into non-natural hosts (Meléndez et al., 1968; Meléndez et al., 1969a; Meléndez et al., 1969b; Meléndez et al., 1972). This discovery was important in the Burkitt's lymphoma (BL) research programme (zur Hausen, 1975, 44-6).

1.2 Medical schemes of causality

There is a substantial medical literature on causality, parallel to that of the philosophy of science.[6] This literature has had an enormous conceptual impact on the manner in which causal relationships have been understood in medical practice. In turn, the changing nature of medical practice has acted to change these medical schemes of causality. A study of causality in medicine that seriously engages with the historical aspects of medical practice must therefore take account of this literature.

Generally, the types of things that these medical theories of causality suggest as characteristic of causal relationships are broadly similar to those found in philosophy. For instance, the requirement for a particular causal factor to be found in most, or all, cases of the disease is common to many of these schemes. We might interpret this as the medical corollary of Humean constant conjunction.[7]

[6] Good medical reviews of this material include Evans, 1976; Lilienfeld and Lilienfeld, 1977; Lilienfeld, 1980; Evans, 1993; Weed, 1995; MacMahon and Trichopoulos, 1996 and Vandenbroucke, 2008.

[7] Despite this similarity, though, there is significant resistance to being 'philosophical', rather than practical, about causality.

I have already suggested that medical theories of causality tend to change with changes in medical practice. Historically, this change is apparent in the types of causal relationships that these medical theories attempt to deal with. This produces three broad classes of medical schemes of causality. Chronologically speaking, the first were those schemes formulated to deal with the demands of infectious disease causation. These sorts of scheme dominated the medical literature on causality between about 1880 and 1960. They were initially developed to account for the causation of bacterial diseases (the HKP),[8] but were later developed to account for viral diseases (Rivers, 1937; Huebner, 1957). They are consistent with the aetiological standpoint, with a clinical disease defined in terms of its causal entity (Carter, 2003). Importantly, and because of this aetiological standpoint, these theories of causality are distinctively *monocausal*. That is, they attempt to demonstrate that a particular agent is the necessary and sufficient cause of a clinical disease.

However, this monocausal stance meant that these theories of causality were poorly suited to the kinds of causal relationships that were discovered in chronic non-infectious diseases (CNID). So new types of causal scheme were formulated to account for these sorts of (more complex) causal relationships, and their use was to become dominant during the 1960s and 1970s. There are two major distinctions that can be drawn between these theories, and the previous infectious disease schemes. First, these schemes are distinctively multifactorial, allowing for the consideration of a great number of causal factors as

---

This can lead to pronouncements with a notable degree of internal tension. For instance, in a discussion of Evans' (Evans, 1978) work on causality we might find the following: 'He offers a practical field guide, eschewing philosophy, in step-by-step procedures for arriving at causation in studies of disease. These procedures essentially realize in practice the triad, in sequence, of Hume's requirements for causality, which can be freely translated as association, time order, and causal direction.' (Susser, 1995, 1125)

[8] Koch's postulates cast a long shadow over causality in medicine. They were first developed to demonstrate the causal relationship between infection with *Mycobacterium tuberculosis* and the development of TB (Carter, 2003. See also reviews by Aronowitz, 2004 and Gillies, 2007). It is well worth taking account of Evans' persuasive arguments for recognising the role of Henle (see Henle, 1938, which is Rosen's translation of Henle's 1840) in forming Koch's opinions about infectious disease causation. I have attempted to do this by adopting Evan's terminology - Henle-Koch postulates, rather than the conventional Koch's postulates - to reflect Henle's influence (Evans, 1976). For more background to the formulation of the HKP, see also Carter's volume of translations of Koch's publications (Carter, 1987) and reviews of same (McGraw, 1989).

The HKP are still used extensively in clinical practice (Grimes, 2006). However, as they depend upon particular technologies to demonstrate cause, the changes in technology that have occurred since their first formulation place the contemporary user in a difficult position. For instance, many have attempted to modify the postulates to take account of non-bacterial illnesses. As an example, Walker et al., 2006, have recently produced a set of 'Koch's postulates' for slow viral diseases and prion disease. I think that this is problematic, as it seems unwise to simply replace Koch's requirement that a causal organism can be grown in 'pure culture' by a requirement that we can, say, express a particular protein in an *in vitro* expression system, without considering the implications for the other parts of the postulates. This concern about incommensurability has not restricted the growth of these modern reformulations. For example, Hackney and Linn, 1979 (HKP for lab research and policy in environmental toxicology); Falkow, 1988 (HKP for molecular microbiology); Hall and Lemoine, 1991 and Wright and Wynford-Thomas, 1991 (HKP for molecular pathology); Brown and Goldstein, 1992 (HKP for cholesterol); Abboud, 1993 and Johnson et al., 1993 (HKP for renal cytokine disease); Fredericks and Relman, 1996 (HKP for molecular microbiology); Fouchier et al., 2003 (HKP for SARS); Walker et al., 2006 (HKP and prion diseases); Inglis, 2007 (an idiosyncratic reformulation of the HKP); Joles, 2008 (HKP for chronic renal disease). There is also a substantial critical literature on the HKP: Stewart, 1968; Evans, 1977 (failure to meet HKP does not necessarily exclude an organism as a cause of disease); Jacomo et al., 2002 (*Bartonella* spp. as exception to HKP); Mazzarello, 2004. There is also a small literature on the HKP as of interest to philosophers of science: Tulodziecki, 2007 being a particularly good example.

being significant in bringing about a single disease. Second, and because of the greater heterogeneity of these CNIDs, their accompanying schemes of cause are not typically couched in strictly methodological terms (as the HKP are). Instead, schemes such as Dorn, 1953; Hammond, 1955; Yerushalmy and Palmer, 1959; Sartwell, 1960; Surgeon General, 1964; Hill, 1965; Wynder, 1966 Yerushalmy, 1966 and Cassel, 1976 tend to suggest general evidential features that pointed to the operation of a cause, rather than a non-causal correlation. Many of these arose from the struggles to demonstrate a causal role for smoking in carcinoma of the lung.

Thirdly, schemes that attempted to find a 'grand unified theory of cause' arose in the 1970s. The authors of these schemes attempted to unify the types of causes that were demonstrated to operate in both infectious diseases and in CNIDs. They make no apparent distinction between cause in these two classes of disease. These schemes fall into two distinct sub-groups. The first are rather general schemes of cause that were intended for fairly casual use in clinical practice. These include Evans, 1976; Evans, 1978; Doll and Peto, 1981; Elwood, 1988; Susser and Rothman, 1988; Kassirer, 1989; Susser, 1991 and MacMahon and Trichopoulos, 1996. These schemes allow the user to apply similar sorts of causal thinking to all kinds of disease. This includes both those which cannot be happily fitted into the acute-chronic categories or, even more cuttingly, those which have moved from one group to the other during the broad historial period that these schemes cover, such as the two examples of viral oncogenesis considered below. This integrated way of thinking about different types of causes also had advantages for understanding causation in infectious diseases, as developments in the understanding of their pathogenesis at about this time began to reveal the important role played by host factors in the aetiology of most infectious diseases. These host factors often presented difficulties for existing schemes of infectious disease causation.

However, these unified theories did not work very well when it came to understanding very complicated causal situations. At the same time, a second group of more sophisticated schemes were also developed that were capable of accounting for highly complex causal relationships. These epidemiological theories of causality stand apart from these more simple, suitable for clinical practice schemes reviewed above. These might be referred to as 'modern epidemiology' causal systems, and there are really five main players. These are the **Sufficient-Component Cause Model** (Rothman, 1974; Rothman, 1976a; Rothman, 1976b; Koopman, 1981; Khoury et al., 1989) which is similar to Mackie's INUS conditions (Mackie, 1974); various medical adaptations of **counterfactual theories of causality** (Rubin, 1974; Hamilton, 1979; Miettinen, 1982; Greenland and Poole, 1988; Rubin, 1990; Robins and Greenland, 1992; Dawid, 2002); **analysis of variance (ANOVA)**; **structural equation modelling** (Spirtes et al., 1993; Pearl, 1995; Pearl, 2000; Robins, 2001) and **Bayesian Networks** (Cornfield, 1976; Pearl, 1993; Crupi et al., 2007). There are also a few instances of attempted syntheses between different species of these causal modelling methods (see, e.g. Greenland and Brumback, 2002).

While there are distinctive differences in the means by which all these different types of theory seek causes, there is one important common factor. All are pluralistic in the types of evidence they take into consideration. Most broadly, they consider evidence arising from observation (of populations,

individuals) and also evidence arising from intervention. For a metascientific account this is excellent evidence in support of a pluralistic account of evidence for causality.

1.3 A note on causal discovery

My historical cases involve the discovery of distinctively causal relations. In what follows, I will use a simple account of causal discovery in order to clarify the manner in which the relationship between causality and evidence changes as such causal claims are constructed.

I suggest that the organisation of medical research programmes change as their central causal claims develop. This change can be seen in the kinds of institutions doing work, the types of research strategies employed, in the kinds of methodologies employed by these researchers, in the nature of the evidence produced by them, and in the ways in which this evidence is employed to strengthen, to alter or to discard these causal claims

Strategies employed early in the history of disease discovery - for example, in attempting to demonstrate that a clinical scenario is real - rather than spurious - or novel - rather than an unusual version of some known disease - will be dissimilar from that causal research found in more mature research programmes. Characteristically, this more mature research would be concerned with attempting to understand the detailed properties of causal entities, and to understand the manner in which they produce particular pathological features of the disease state. This concern with detail is sociologically apparent in the extreme specialisation typical of mature medical research programmes. Researchers may spend entire careers working on one aspect of one cause of one disease without arousing particular surprise. Their understanding of causality, and the types of evidence they think about, will be very different from that of the early researcher in the field, attempting to discover if the (diverse) clinical features of a particular individual represent some kind of new disease. Similarly, their causal methodology may well be quite different from the specialist researcher working on a different area of research into the same disease.

This shifting nature of the kinds of questions and answers encountered in these long - and diverse - research programmes thus presents us with an historical problem: we cannot understand causality or evidence for that causality monolithically. Instead, in order to deal with causation in medicine properly, we will need to treat it relative to the degree of development of the research program into the condition. In order to do this, I have employed a simple account of causal discovery to make these distinctions between different phases of research. This account characterises different phases of research on the grounds of the type of causal claims sought as the basis for forming an appropriate distinctions between historical periods.

It is based upon Kuhn's model of discovery (Kuhn, 1962). While this was developed around historical examples from the physical sciences (Priestley's discovery of oxygen (or dephlogisticated air), Herschel's discovery of Uranus and Roentgen's discovery of x-rays), it seems to be highly compatible with causal discovery in medicine. Briefly, Kuhn argues that discoveries are processes that are extended across time, persons and places. This extension grants discoveries a describable internal history:

> ...discoveries have a proper internal history as well as prehistory and a posthistory.

> Furthermore, within the rather vaguely delimited interval of internal history, there is no single moment or day which the historian, however complete his data, can identify as the point at which the discovery was made. Often, when several individuals are involved, it is even impossible unequivocally to identify any one of them as the discoverer
>
> (Kuhn, 1962, 763)

Now this extension of the history of discovery appears right for these medical cases of causal discovery too. For example, the discovery of the manner in which infection with EBV causes BL took roughly 20 years to firmly establish, and involved a great many researchers. This suggests that these causal discoveries in medicine have a rich internal history. Kuhn suggests that this internal history of discovery can be divided into three, characterisable, stages. First comes the production and recognition of an anomaly. Discoveries, he claims, begin with the awareness that a particular observation or experimental finding is, in some way, remarkable. As he suggests, the production of such anomalies suggest that all is not as it should be, and such anomalies are therefore indicative of a tension between an existing theory and an observation. Anomalies are 'nature's failure to conform entirely to expectation.' (Kuhn, 1962, 762).

However, it is not the case that these anomalies are necessarily easy to recognise. In fact, Kuhn suggests that for an anomaly to be meaningfully recognised as such, two prerequisites must be in place. The first of these is that the finder of the anomaly must recognise the anomaly as something striking that requires further investigation:[9]

> The first... is the individual skill, wit, or genius to recognize that something has gone wrong in ways that may prove consequential. Not any and every scientist would have noted that no unrecorded star should be so large, that the screen ought not have glowed, that nitrous air should not have supported life.
>
> (Kuhn, 1962, 763)

The second of these prerequisites is that there must be appropriate theories and technologies available, such that the anomaly can be both produced and recognised as anomalous:

> Whatever the level of genius available to observe them, anomalies do not emerge from the normal course of scientific research until both instruments and concepts have developed sufficiently to make their emergence likely and to make the anomaly which results recognizable as a violation of expectation.
>
> (Kuhn, 1962, 763)

This production and recognition of an anomaly is merely the start of the internal history of a discovery. Kuhn's second stage is the formulation of this anomaly and making it lawlike. This can be characterised as:

[9] A good example of this is Burkitt's very early description of BL, which I discuss in sections 2.2.3.

> . . . a more or less extended period during which the individual and often many members of his group struggle to make the anomaly lawlike. Invariably that period demands additional observation or experimentation as well as repeated cogitation. While it continues scientists repeatedly revise their expectations, usually their instrumental standards, and sometimes their most fundamental theories as well.
>
> (Kuhn, 1962, 763)

Finally, he suggests that the process of discovery is not complete until the implications of the new discovery are investigated. Rather than suggesting that discovery is typically just about finding new facts that can be added, in an aggregative fashion, to the store of total knowledge. Kuhn argues that important new discoveries fundamentally change scientific practice, making them assume much greater importance than they would have as simple facts:

> Very often these transformations in the established techniques of scientific practice prove even more important than the incremental knowledge provided by the discovery itself.
>
> (Kuhn, 1962, 763)

In what follows, I employ a slightly modified version of Kuhn's model of discovery that is based upon cases of causal discovery in the history of medicine. Unlike Kuhn, this is an account that is therefore intended to deal with the discovery of the detailed operation of causes, rather than one that begins with the discovery of existence claims about entities (as with Kuhn's examples).

First, I think that something happens to causal anomalies before they become really suitable for the kinds of collaborative investigations of Kuhn's second stage. In other words, I want to tell a fuller story about the internal history of anomaly recognition than Kuhn. Not only do researchers discover an anomaly, but their initial investigations seek to determine the kinds of ways that it may come about. I thus break Kuhn's first stage, of anomaly production and recognition, into two distinct stages. I suggest that researchers first **suspect** that a novel type of cause is responsible for some kind of anomaly. However, they then perform early investigations to detect over what kinds of things this cause operates. I term this kind of research **domain finding**. Perhaps this division between types of research activity is related to Kuhn's two prerequisites of discovery, with the individual 'wit' of a researcher underwriting the process of suspicion, while the business of domain finding is the causal corollary to the 'violation of expectation' from Kuhn's account.

We might define causal domains as the result of a first-pass classification of possible, likely causes. I suggest that a newly discovered anomaly must be placed in a broad causal category. To illustrate, does an anomaly result from the actions of some kind of infectious cause? Or is it produced by a neoplastic cause? A congenital abnormality? Broadly, deciding the particular domain of a cause is a problem of classifying possible causes for a particular phenomenon.

Two related example drawn from medical practice serve to illustrate both that this kind of behaviour occurs, and plays a significant role in learning about causes. The first, rather well described, is that of

seeking a differential diagnosis for a particular condition (see, for example, Bouchier et al. (1996) as just one example of entire textbooks devoted to this topic).

The second, while being a key feature of medical practice, is rather less well described.[10] This is classification by an *aide-mémoire* known colloquially as the 'surgical sieve'. This kind of classificatory behaviour consists of generating a list of possible causes by reference to a learned set of possible types of cause. So we might use a surgical sieve 'consisting of normal, developmental, trauma, infection, neoplasia, inflammation, vascular, metabolic' (Toms et al., 2006, 668) to provide us with the possible types of disease process as classificatory categories. We can then use these broad pathological categories to generate the likely possible diagnoses operating in a particular instance - that is, to form a differential diagnosis. For instance, Stell (1977, 265) classifies the causes of epistaxis (bleeding from the nose as follows). First, epistaxis may result from either congenital abnormality or acquired insult. These acquired causes can either be infections (acute or chronic), tumours, trauma, disorders of coagulation, or idiopathic (i.e. no cause known). These classifiers, provided by the surgical sieve, are general, non-specific, and predicated more on plausibility than anything else, then act as containers for much more specific diagnoses evoked when producing a differential diagnosis. Stell (1977, 265), for example, gives the following specific suggestions as examples of chronic infections likely to cause epistaxis: infection with syphilis, infection with tuberculosis, infection with *Klebsiella ozaenae* (ozaena) and so on. In contrast then, these possible causes are local, deal with specific possible causes, and requires thinking about the specific manner in which a particular effect comes about.

The causal domain is therefore a manifestation of the kind of simple classificatory behaviour represented by the surgical sieve.[11] Why is it important? First, it plays a vital sociological role. For instance, which investigators should investigate a newly discovered x? Second, there is a closely related methodological point: which research practices should be used to investigate x, and why? The researchers and research practices involved in investigating an anomaly that is suspected to arise from the actions of an infectious agent will be quite different from those used to investigate something with a presumptively neoplastic cause.

I also differ slightly from Kuhn's position when it comes to the later stages of discovery. The first of these is the process of **mechanism construction**. This phase is where researchers develop an understanding of the specific causal processes that account for the observed anomaly. Finally, researchers assess this newly developed putative causal mechanism by **applying** it in disease populations. Perhaps the full realisation of this stage also consists of some of Kuhn's third stage of discovery. This application to disease populations therefore both tests the new causal mechanism, and aims to discover where and when it occurs. Perhaps this slightly different distinction results from the role of laws of nature in medicine. Metascientifically, it is not the case that the discovery of a causal relationship in medicine is

[10] However, it really is very well known in medical practice, and familiar enough to form the basis of a humorous piece in a festive edition of a major medical journal without explanation (Turmezei, 2009). Briefly, this is a surgical sieve represented in the form of Beck's renowned mapping system for the London Underground network.

[11] I avoid the term: it sounds more like equipment than anything else. In any case, there is very often nothing surgical about the kinds of relationships I want to use it to characterise.

described as the discovery of a causal (or indeed, any other) sort of law.

So causal discovery in medicine is characterisable into four stages, and I will use them chronologically in the case histories. A further characteristic of each of these stages is their evidential pluralism. At no particular point is any single type of evidence or investigation necessary or sufficient to define the discovery or characterisation of a causal process in medicine.[12] As a related issue, these stages of causal discovery seem to operate somewhat recursively. First, a novel causal mechanism is suspected to be responsible for a particular state, *E*. Next, the likely domain of possible causes, *C*, are sought. After this, a mechanism linking *C* and *E* is constructed. At this stage, the overall *C*-*E* causal mechanism is broken down into a series of smaller causal processes, *C*-*C1*-*C2*...*Cn*-*E*. Each of these smaller links is then investigated in the same way. Their existence is first suspected, and their domain sought, before they, in turn are related to the operation of a mechanism.[13]

The cases of viral oncogenesis presented below have formed the evidence base for this account of causal discovery. It is also illustrated by the other histories of causal discovery mentioned in section 4.5. However, owing to the particular contingent difficulties of establishing the existence and nature of causal relationships in these cases, I am reluctant to defend this as a general account of causal discovery in medicine. In particular, I suspect that in the many cases where discovering causes was much less tricky than it was in these instances, the characteristic differences between these stages will not be apparent in the historical sources available. After all, if suspicion, domain-finding and mechanism construction all occur as part of the research conducted for a single paper, this characterisation will not be apparent. In what follows, I will therefore use it simply to clarify these histories of viral oncogenesis.

[12] c.f. also Bechtel and Abrahamsen, 2005, 426

[13] An historical example of this recursion is provided in section 2.4.

Chapter 2

Epstein-Barr Virus and Burkitt's Lymphoma

2.1 Introduction

2.1.1 Chapter outline

Burkitt's lymphoma (BL) is a tumour that is caused by infection with Epstein-Barr virus (EBV). In this chapter, I outline the manner in which this causal relationship was discovered. I will begin with a brief note on the contemporary understanding of BL and EBV. I will then move back to the earliest parts of the story in section 2.2 to discuss the discovery of BL as a clinical syndrome. Section 2.3 will then detail the initial investigations of the pathology and basic epidemiology of this tumour syndrome in the field. During this period, researchers used a variety of evidence and techniques to attempt to discover what sort of causal mechanism[1] could account for the unusual features of this tumour syndrome. The important conclusion arising from these investigations was the suspicion that the disease was a malignant one, possibly caused by a virus. The next step was the construction of a causal mechanism, linking the development of the tumour with a viral infection. I will therefore shift my focus to the laboratory in section 2.4, with a review of the investigations that led to the discovery of EBV, and the construction of a mechanistic model of its involvement in the pathogenesis of BL. Finally, in section 2.5, I will show how this mechanistic model was employed to investigate populations at risk of BL in the field.

This story of the linked discoveries of BL, EBV and the BL-EBV causal mechanism, is notable in a number of respects. It was the first example of viral oncogenesis discovered in humans. As schemes for demonstrating causality had historically treated tumours and infectious diseases rather differently (see section 1.2), this case required researchers to think about causality in novel ways. For an infectious agent to cause cancer, the nature of causal claims available to medical researchers had to undergo fundamental adjustment. Not only did these causal claims have to be unusually transparent and robust, but researchers also had to appraise the usually tacit assumptions underlying the development of causal claims in medicine. Instead of remaining in the background, they were drawn out, critically examined and in some cases reformulated as part of the process of discovery. For the time being, I suggest that the most important of these relates to evidential pluralism. Not only did researchers look for constitutive

[1] In this and the next chapter, I will use the term *mechanism* in a philosophically agnostic fashion. By a mechanism, I mean to refer simply to some sort of underlying series of linked causal processes. I will return to mechanisms in chapter 5 to give a more sophisticated reading

or supportive mechanistic evidence[2], but they also reasoned by analogy with the mechanisms of other disease processes. Multiple types of evidence are often assumed to be characteristic of truly causal processes in medicine. But it is only in this sort of highly contentious causal situation where this assumption is examined. In many ways, a pluralistic approach to evidence is the defining characteristic of this historical case. This evidential pluralism operates across the four stages of causal discovery discussed in section 1.3 (suspicion, domain-finding, mechanism construction and application). More interesting, it operates across them in a recursive manner. So not only do we find these steps during the process of causal discovery linking the clinical syndrome of BL to infection with EBV, but we also see similar multi-step processes occurring at interim points. As an example, in section 2.4, I show just such an example of nested causal discovery in EBV research. I will go on in later chapters to discuss this process of causal discovery in terms of the Russo-Williamson thesis (Russo and Williamson, 2007).

All this means that the inner workings of causal discovery are unusually prominent in the BL-EBV literature. In particular, both the clinical discovery of the lymphoma syndrome and the early laboratory work leading to the detection of EBV are unusually explicit in their examination of the assumptions that usually remain tacit in medical research. These are therefore the sections of the story that I will focus on. A fuller picture of events as they came about can be found in a number of historical reviews, such as Epstein, 2001a; Epstein, 2005; zur Hausen, 2006 and Stebbing and Bower, 2009. BL is exceptional for other reasons too, as it played a paradigmatic role in cancer biology during the late twentieth century. Likewise, much of our understanding of the detailed functions of both RNA and DNA viruses, including our knowledge of viral latency, has derived from work on EBV. More broadly, our understanding of the functioning of gene promoters and of cell-cycle regulation arose in large part from the early work on B-cell immortalization by EBV. Thus, the BL-EBV story has held greater importance to medical researchers than perhaps it deserves as a purely clinical problem (Galloway, 1989, 464).

2.1.2 Burkitt's Lymphoma

So what is BL? It is an extranodal B-cell lymphoma of mature form (Harris et al., 1999, 3837; 3844ff). *Lymphoma* describes a solid blood-cell tumour, particularly of lymphocytes or lymphocyte precursors.[3] *B-cell* refers to B-lymphocytes, and *mature form* suggests that the tumour cells resemble normal B-lymphocytes, which are the immune cells responsible for mounting antibody attacks against foreign agents. Finally, *extranodal* suggests that the disease tends to form solid tumours outside lymph node structures.

Clinically, the tumour usually presents as a large mass in the bones of the jaw, often involving more

[2] As with the use of the term mechanism, I do not mean this to carry any baggage for the time being. I will develop the term *mechanistic evidence* in chapter 6

[3] Leukaemia, in contrast, refers to blood-cell tumours which exist in a distributed form in the blood and bone marrow, rather than constituting a solid tumour. There are a range of intermediate forms between lymphomas and leukaemias (Magrath, 2002, 2377). For instance, some patients with lymphomas may begin to develop a leukaemic stage with cancer cells circulating in the blood. This happens in some cases of BL, meaning that strictly speaking researchers now construe it existing as a spectrum of disease, with BL as one extreme, Large B-Cell Lymphoma (LBCL) at the other and with an intermediate category of Burkitt-Like Lymphoma (BLL) between the two. (Harris et al., 1999, 3845; Magrath, 2002, 2379; 2391). But the terms leukaemia and lymphoma are descriptive enough to sustain the following discussion.

than one quadrant[4] with further tumours in the abdomen (Ziegler, 1981, 738; Magrath, 2002, 2390). It is highly malignant, possibly the most rapidly growing tumour of humans (Iversen et al., 1974; Ziegler, 1981, 738). It has a characteristic age incidence, with almost all cases occurring between 2 and 16 years, peaking at 7 years. There is a male:female ratio of about 2:1 (Ziegler, 1981, 738). Globally, it constitutes about 5% of all childhood malignancy (Magrath, 2002, 2378).

Histologically, the cells are relatively monomorphic (i.e. all the tumour cells appear similar), with 2-5 nucleoli and an 'open' nuclear chromatin pattern (Magrath, 2002, 2378). The tumour is said to display a 'starry sky' pattern, owing to the presence of tingible body macrophages, although this appearance is not specific for BL (Magrath, 2002, 2379). The cells have a germinal centre phenotype (CD38-positive) and 90% express IgM surface immunoglobulin with κ or λ light chains. The cells also express typical B-cell antigens.[5]

There are three forms of the disease.[6] Most cases occur in tropical Africa, and are known as *endemic* BL. Outside this area exists a *sporadic* form of the disease. Sporadic BL affects slightly older children, who more often develop abdominal tumours (Ziegler, 1981, 738). The two forms also have different cytogenetics, with a degree of variability in the types of chromosomal translocations found in them. However, the tumour cells, response to treatment and patient prognosis are similar (Ziegler, 1977; Ziegler, 1981, 738). There is also an AIDS- or immunosuppression-associated form of the disease (Hecht and Aster, 2000, 3708). While aware of this distinction between types, I will speak exclusively of the endemic form in this chapter.

2.1.3 Epstein-Barr Virus

EBV is a B-lymphotropic γ-herpesvirus. It infects B-cells, and is capable of immortalizing them (Kieff and Rickinson, 2001, 2511-5). The virus is highly prevalent in adult human populations worldwide, with up to 90% of adults displaying evidence of infection (Henle et al., 1969). Detailed seroepidemiological studies have, however, revealed that the age of first infection differs greatly between populations. EBV infects children in the developing world much earlier than in developed countries (Evans and Niederman, 1989). This difference in age at primary infection has been explained in terms of differential mechanisms of transmission between populations. As EBV is transmitted via buccal fluid (Gerber et al., 1972), any activities that share saliva can cause the transmission of the virus. For example, pre-chewing of food by adults as a weaning practice in developing countries is often held as a cause of childhood infections with EBV.[7]

Clinically, primary infection is usually asymptomatic, although a minority of adolescents and adults develop infectious mononucleosis (IM). This is a non-malignant syndrome of fever, cervical lym-

[4] Right and left hand sides of the maxilla (upper jaw) and mandible (lower jaw)

[5] CD10, CD19, CD20, CD22, CD 77 and CD79a positive (Magrath, 2002, 2378)

[6] An excellent comparison between the forms is Thorley-Lawson and Allday, 2008, 914: box 1

[7] However, the precise mechanism by which EBV reaches the saliva remains controversial (Allday and Crawford, 1988; Niedobitek and Young, 1994; Rickinson and Kieff, 2001). Some suggest that the virus, in addition to infecting B-lymphocytes, is also active in epithelial cells of the oral mucosa, which are then shed in saliva (Sixbey et al., 1984). Others argue that the virus reaches the saliva through a lytic form of infection occurring in mucosal lymphocytes (Weiss and Movahed, 1989; Tao et al., 1995; Karajannis et al., 1997).

phadenopathy and sore throat, followed by a lengthy period of fatigue and malaise. Research suggests that infection with the virus relatively late in life predisposes to developing IM, and this suggestion is born out by the very low incidence of IM in the developing world, where the average age at primary EBV infection is very much lower than in the West. While IM will play a role in this chapter (see section 2.4.3), a much greater one will be played by diseases associated with EBV re-activation. The virus can remain dormant in B-cells (Kieff and Rickinson, 2001, 2515), often for very many years. Rarely, long after initial infection, the virus become active again in the B-cell population or in other cells, resulting in a range of malignant human diseases.[8] These include BL, nasopharyngeal carcinoma (de Schryver et al., 1969; zur Hausen et al., 1970), gastric cancers (Herath and Chetty, 2008), Hodgkin's lymphoma (Pallesen et al., 1993), childhood leimyosarcomas (Lee et al., 1995; McClain et al., 1995), T-cell and NK-cell lymphoma (Pallesen et al., 1993) and the premalignant condition oral hairy leukoplakia (Komatsu et al., 2005). There are also a number of EBV-associated tumours found specifically in HIV/AIDS patients (MacMahon et al., 1991) or in individuals with other forms of immunosuppression (Hopwood and Crawford, 2000). EBV has also been tentatively suggested to be a possible cause of some chronic non-malignant diseases, including multiple sclerosis (Pender, 2009) and chronic fatigue syndrome (Glaser et al., 2005).

The world-wide distribution of EBV and the wide range of clinical manifestations of EBV-associated disease presented major difficulties for researchers attempting to show that the virus caused BL, as I will discuss in section 4.4.1.

2.2 Suspicion: BL as a tumour syndrome

2.2.1 The clinical discovery of BL

Burkitt's first description of the lymphoma syndrome made the following points.[9] First, while malignant tumours of the jaw were thought to be rare in children, they appeared to be unusually common in parts of Uganda. In fact, in the region of the Mulago hospital on the outskirts of Kampala, they were by far the commonest malignant disease of childhood (Burkitt, 1958, 218; 223). Based on 38 cases occurring during the period 1951-8, he described an apparently novel jaw tumour syndrome (Burkitt, 1958, 218). This tumour was multifocal, with histologically similar material found both in the bones of the jaw and at other anatomical sites, including within the abdomen (Burkitt, 1958, 219; 221-3). He then gave a histological description of the tumour,[10] before speculating about its origins and nature:

> ...the clinical picture of this sarcoma is easily recognized, [but] its site of origin and nature remain obscure...It may well be that we have in Uganda a syndrome of diverse characteristics...which has attracted attention owing to its very striking clinical features.

[8] For a review see Deyrup, 2008

[9] A note on terminology. The term *Burkitt's lymphoma* dates from the 1962 UICC conference on reticuloendothelial tumours in Africa (Glemser, 1971, 124). Despite this anachronism I will use the term from the outset for the sake of clarity.

[10] '[the tumour is]...highly malignant...with a large open vesicular nucleus, usually showing two prominent nucleoli. The cytoplasm is scanty and mildly eosinophilic. These cells occur singly or in clusters, often in rounded, otherwise clear spaces surrounded by masses of small, dark, round cells with dark compact hyperchromatic nuclei and scanty cytoplasm and strongly resembling lymphocytes; in some cases the tumour resembles a lymphosarcoma.' (Burkitt, 1958, 220).

> Still less is it intended to speculate as to why this sarcoma should occur so frequently here. So far no common factor has been recognized which might have a bearing on the aetiology.
>
> (Burkitt, 1958, 221-3).

So Burkitt here described the clinical features of an unusual, possibly novel, tumour syndrome, with high incidence in Uganda. His description, at this stage, did not concern the details of any underlying pathological mechanism. However, as the tumour appeared to strike at many different parts of the body simultaneously, including multiple sites within the jaw and within the abdomen, this clinical description itself implied some kind of aetiological novelty - most primary tumours[11] do not behave like this.

For these phenomenological features of the tumour syndrome to be accounted for by a malignant process, a novel mechanism of simultaneous, multifocal oncogenesis occurring in both jaws and abdomen of each of these subjects would be required. The alternative would be equally challenging: if the primary tumour began in either the jaws or abdomen, and metastasised to the other sites, a novel mechanism of simultaneous, specific and multifocal metastasis would be needed.[12] For the time being, Burkitt suggested neither. Instead, he stressed the novelty of the clinical picture *qua* clinical picture.

At this juncture it seems worthwhile to give a brief biographical sketch of Burkitt, together with some historical background to this first description of BL. I will return to the lymphoma syndrome presently.

2.2.2 Denis Burkitt

Denis Burkitt (1911-1993) was born in County Fermanagh, Northern Ireland. After initially training as an engineer, he read medicine at Trinity College Dublin, graduating in 1935. After completing positions as a house surgeon, he successfully sat the examinations for the Fellowship of the Royal College of Surgeons of Edinburgh in 1938, before sailing as a ship's surgeon to China. While on this voyage, he made a decision to continue his surgical career in the developing world (Epstein and Eastwood, 1995, 91). During the Second World War, he applied to join the Royal Army Medical Corps and, at his second attempt (Glemser, 1971, 36) was posted to Uganda in 1943. Following demobilisation in 1946, he applied to the then Colonial Office in order to continue his work in Uganda, and was briefly posted to Lira, in the Lango district before being transferred to the Mulago hospital on the outskirts of Kampala in 1948 (Epstein and Eastwood, 1995, 92), where he was to remain until he retired from surgery in 1964. A period of working for the Medical Research Council then followed, initially in Uganda, and then in London from 1966-76 (Heaton, 1993, 951).

Studying Burkitt's bibliography (available ff. Epstein and Eastwood, 1995) the range of his early work - including publications on vascular surgery, urology, orthopaedics and general surgery - is remarkable. This is perhaps not so surprising given his likely range of clinical duties at the Mulago hospital. Located on the outskirts of the Ugandan capital, and now a flourishing university hospital, at the time

[11] Primary tumours refer to those found at the original site of the tumour. Secondary tumours, on the other hand, are those that have spread - metastasised - from a primary tumour at a remote body site

[12] Most tumours metastasis by either spread in the blood, in the lymphatic system, or by direct contact - for example within the abdomen. None of these mechanisms appears a good fit with the clinical details of this tumour

of Burkitt's appointment, it 'consisted of thatched, single-storey huts housing the wards together with rudimentary ancillary facilities' (Epstein and Eastwood, 1995, 92). It was at this institution that he was to describe his eponymous tumour. How did a general surgeon - without specialist experience of either paediatrics or oncology - come to describe this disease?

2.2.3 Background to the clinical discovery of BL

Convention dictates that Burkitt first encountered the disease in 1957. Burkitt was called by his colleague Hugh Trowell (1904-1989) to see a 5 year old boy named 'Africa' in Ward One of the Mulago Hospital (Glemser, 1971, 49). The boy's symptoms were strikingly unusual, but did not cause Burkitt to suspect a significant anomaly in need of investigation:

> His face was massively swollen, with bizarre lesions involving both sides of his upper and lower jaws. I had never seen anything like it. The teeth were loose and the features grossly distorted. If a single jaw quadrant had been involved, I might have considered it to be an infective process such as osteomyelitis, but not with all four quadrants affected. This unusual distribution also seemed to rule out any form of neoplasia. Results of the biopsy had suggested some form of granuloma. I was totally baffled, but photographed the child and considered this to be another of the curiosities one had to become accustomed to seeing from time to time in Africa.
>
> (Burkitt, 1983, 1777).

However, it appears that this wasn't the first time that Burkitt had been made aware of unusual tumours of the jaw in children in Uganda. For one, these tumours had been noted by other medical offers at Mulago since at least 1948 (Glemser, 1971, 49). Further, a record of a clinical meeting in Mulago entitled 'Tumours of the Jaw' on the 29th October, 1955 exists, at which Burkitt was present (Singh, 1955). In this note, the higher incidence of jaw tumours at Mulago than in either Europe or America was noted (Singh, 1955, 70). Furthermore, the histology of at least ten of these tumours were suggestive of the Burkitt type.[13] Burkitt is mentioned in the discussion of these 'small round-cell sarcomas'. This suggests that Burkitt was already somewhat acquainted with these unusual jaw tumours before meeting his 'first' case.

Some evidence also suggests that Trowell had previously encountered the tumour:

> Once, in 1935, when he [Trowell] was on safari to the islands of Lake Victoria, inspecting the population for evidence of sleeping sickness, he came across a little boy with "lumps" on the jaws. 'I can't remember much about it except that I thought it was a peculiar complaint, and I told the boy, I can't treat you here, you must come back to Mulago with me. I didn't know what the disease was, because it wasn't one lump, which might be called

[13] They are recorded as 'lymphosarcomas... mainly in the younger patients... [which] were small round-celled growths without rosettes' (Singh, 1955, 70). As I will discuss, there was no clear referent for Burkitt-type tumours before Burkitt's description. The terms *lymphosarcoma* and *reticulum-cell sarcoma* are often considered to encapsulate the majority of tumours that would now be described as Burkitt's lymphoma or Burkitt-like lymphoma.

> an ordinary cancer, but several lumps. Vaguely, at the back of my mind, was the knowledge that the children out here had these lumps: one encountered them from time to time, one would ask surgeons to see them but nobody could throw much light on the trouble.'
>
> (Glemser, 1971, 60).

E.H. Williams, who would later publish extensively on the epidemiology of BL (see e.g. Burkitt et al., 1963; de-Thé et al., 1978) also recounted seeing several cases of the syndrome over the years (Glemser, 1971, 102). Others did too. For instance, Sir Albert Cook's 1901 records from Mengo hospital in Uganda mention similar jaw tumours (Hutt, 1981, 762). There are further reports noting the high frequency of jaw tumours in Uganda, dating from at least 1897 (Davies et al., 1964, 339 and appendix 1).

So it is clear that this phenomena was fairly wide-spread, and that many individuals were in some way aware of it before Burkitt. However, it was Burkitt who meaningfully recognised the strangeness of these jaw tumours.[14] His curiosity was pricked when he saw a child with similar symptoms to Africa a few weeks later while visiting the district hospital in Jinja, on the northern shores of Victoria Nyanza, fifty miles from Kampala (Glemser, 1971, 50):

> A few weeks later, I was doing a ward-round in another hospital which I visited regularly... my attention was attracted by a child with a swollen and distorted face... to my surprise, recognized precisely the same features as I had observed in the Mulago Hospital ward a few weeks previously. My interest was rivetted immediately. A curiosity can occur once, but two cases indicated more than a curiosity.
>
> (Burkitt, 1983, 1777-8)

2.2.4 Two cases become a syndrome

Burkitt began to believe that these two patients were not isolated oddities (Burkitt, 1983, 1778). I suggest that it was at this point that Burkitt began to suspect the possibility of a distinct causal mechanism underlying this tumour syndrome. More than just jaw tumours, other features of these cases, such as abdominal tumours, were suggestive of fundamental, generalizable similarities. This association between jaw and abdominal tumours had been previously noted at Mulago. Burkitt's examination of the hospital records, spurred by these two cases, led to a growing recognition of this as a significant unifying feature of the syndrome. He recounted how he began formulating his description of a clinical syndrome based on these cases:

> I went retrospectively through the records of all children suffering from malignant tumors, and the majority were found to be in certain particular sites, including the jaws, orbit and eye, kidneys, adrenals, and ovaries and, occasionally, in the liver, testicles, thyroid gland, and long bones... A characteristic feature was the usual absence of peripheral lymph-node involvement. Two points became abundantly clear: first, these various tumors tended

[14] The nature of this anomaly recognition (c.f. Kuhn, 1962) is something that I address in sections 1.3 and 6.3.1

> to be associated with one another in individual patients; and second, since the jaws were not always involved, tumors there could not be the primary lesion. The recognized associations between what were previously assumed to be totally different diseases prompted the conclusion that whatever the nature of these tumors, they must be part of the same disease process, a conclusion that was strengthened by the similar age distribution regardless of where the tumor was cited.
>
> (Burkitt, 1983, 1778)

There was also histopathological consistency, not only between both tumours in each patient, but between the tumours of different patients:

> C. T. O'Conor and J. N. P. Davies went retrospectively through histologic slides of cancers in children, and came to the conclusion that all of these tumors were not only histologically identical, but also were some form of lymphoma.[15] Consequently, they altered previously made diagnoses and what had emerged as a clinical syndrome became a pathologic entity.
>
> (Burkitt, 1983, 1778)

So by the time these first two cases were reported, the clinical features, histological consistency and the suspicion of a clinical cohort with a characteristic age incidence were already present. These initial findings were presented at a staff meeting at the Mulago hospital in 1957 (Glemser, 1971, 68-9), before a more formal presentation at the annual meeting of the East African Association of Surgeons in Kampala, during January 1958 (Glemser, 1971, 106). The 1958 paper developed from these two presentations. Remarkably, however, it was not widely read, despite its apparent importance as a novel description of an apparently common condition. Burkitt is most clear upon the reasons for this failure:

> ...I wrote my first paper describing this tumor and, as a surgeon, considered it an opportunity to get an article into the British Journal of Surgery. In retrospect, it was the wrong journal in which to report work of this nature, and it received minimal attention.
>
> (Burkitt, 1983, 1779)

Happily, this indifference was short-lived, and subsequent publications were much more successful. These later findings, though, represented a significant shift in the direction of BL research. The first BL publications largely attempted to make the case for the possible existence of a distinct causal mechanism underlying the observed clinical features of BL. Later papers, however, in the main assume the existence of some sort of causal mechanism, and focus on demonstrating and exploring its features. Notable are the attempts made to demonstrate the novelty of this mechanism, and the attempts made to determine the causal domain within which this new mechanism operated.

[15] Discussed in O'Conor and Davies, 1960

2.3 Domain finding: BL as a geographical disease

This phase of research therefore comprised two related branches of investigation. First, researchers demonstrated that the suspected mechanism was a novel one. This involved giving a detailed, unified clinical description of the disease, with the intention of ensuring that this syndrome was not, for instance, an unusual manifestation of another, known disease. One characteristic part of this process was the distinction drawn between the lymphoma syndrome and leukaemia, as I discuss in section 2.3.2. Second, researchers attempted to determine the nature of the domain of the suspected causal mechanism. This began with the demonstration of the malignant nature of the syndrome. Then, the character of research altered towards attempting to discover the domain of the likely causes of this malignancy. In large part this focused on the possible infectious nature of the tumour, as suggested by its unusual epidemiology. At the end of this phase, though, there was no detailed causal mechanism capable of explaining the occurrence of BL. Instead, what was understood was the likely nature of this, as yet undiscovered, mechanism.

2.3.1 Building a clinical picture

The second, third and fourth papers published on the lymphoma syndrome detailed the clinical and epidemiologic features of the tumor (Burkitt and O'Conor, 1961), and its pathological and histopathological features (O'Conor and Davies, 1960; O'Conor, 1961). Unlike their predecessor, they were to arouse global interest in the tumour (Burkitt, 1983, 1779).

Burkitt and O'Conor, 1961 described a review of the lymphoma syndrome based on 106 cases aged from birth to 14 years living in the Mengo district of Uganda. The clinical features were similar to those previously documented. The authors found that the disease had a peak incidence at ages 3-8 years (Burkitt and O'Conor, 1961, 262), and recorded that abdominal tumors seemed to be present in all cases with jaw tumours. Furthermore, the authors claimed that the syndrome had a generalized distribution in an East-West belt across Equatorial Africa, with localized areas within this belt having a much higher BL incidence than others. When combined with the observation that tribal and racial factors did not alter the disease incidence, the authors therefore concluded that the syndrome was likely to be caused by the influence of environmental factors.

O'Conor and Davies, 1960 confirmed that the jaw and abdominal tumours seen in the lymphoma syndrome were histologically indistinguishable, poorly differentiated lymphosarcomatous lymphomas (O'Conor and Davies, 1960, 531-4). O'Conor, 1961 reinforced the other descriptions of clinical features and offered a histological characterisation of the tumour, reporting that 95% (101 of the 106 cases) were of similar histological appearance,[16] and went on to suggest that this histological homogeneity was evidence for these clinical cases being instances of a unified clinicopathological entity:

> As has already been inferred from the histological description, there is a closer morphological and apparent cytogenetic relationship between the tumors of this entire series than one would expect to encounter in such a large group of lymphomas, so that they seem to form a histological entity as well as a clinical syndrome. If the purpose of classification is

[16] In fact, 83% were 'poorly differentiated lymphocytic', 10% stem-cell type and 2% mixed-cell type (O'Conor, 1961, 277).

> primarily to indicate a probable course or prognosis, there would hardly be justification for division within these 106 cases...
>
> (O'Conor, 1961, 277)

Before long, the multicentric nature of the tumour, and its strong preference for unusual anatomical sites[17] became, rather than a diagnostic puzzle, a characteristic feature of the syndrome (Burkitt, 1962a, 232-3; Burkitt, 1962d, 217; Burkitt, 1962f, 71). This provided further evidence for the unified and distinctive nature of the disease. This process of unification was exceptionally clinically important:

> It may be wondered how this tumour syndrome has passed unnoticed until the last few years. Individual trees had called for comment, but the wood had apparently not been seen. Rapidly growing tumours had been observed in the jaws of children, but had not been recognized as a specific entity and part of a multicentric tumour syndrome. Also the frequent occurrence of ovarian lymphosarcoma had been reported from the Cameroons (Capponi, 1953). It was the recognition that many unusual and hitherto unconnected tumours were, in fact, but different manifestations of one tumour syndrome that initiated investigation of the problem as a whole.
>
> (Burkitt, 1962f, 75-6)

The features of the disease were refined by three early reviews of the clinical aspects of the lymphoma syndrome (Burkitt, 1962a; Burkitt, 1962d; Burkitt, 1962f), each of which spelled out the malignant nature of the disease. Burkitt's other three papers of 1962 (Burkitt, 1962b; Burkitt, 1962c; Burkitt, 1962e), though, focus on finding the domain of the likely causes of this malignancy, largely by investigating the geographical features of the disease epidemiology. What were these geographical features?

First, that BL is a common and distinctive syndrome of children in Africa, as opposed to a disease of children of African descent. While Burkitt's initial cases did not include any Caucasian children, and only one each of Asian descent and of mixed race, this was to be expected given the relative proportions of different races in the catchment of the Mulago hospital. Burkitt was most careful to stress that the syndrome was not specific to any one racial, tribal or social group and was, instead, a disease of all children living in Africa, with an apparent bias towards black Africans purely because of their majority status within the population.[18]

Second, that this disease was not new, being instead locally and anecdotally reported long before the recognition that it was a distinct clinical entity (Burkitt, 1962a, 232; Burkitt, 1962d, 211; Burkitt, 1962f, 71). When the clinical prominence of the lymphoma syndrome is considered, it is surprising that Burkitt was the first to describe it as a common clinical entity. Some justification for this may be found in the early publications on the subject:

[17] Not only does BL tend to attack parts of the body usually spared by malignant disease, but it also spares other sites which are usually considered highly susceptible, such as the axial skeleton, lymph nodes and the lungs.

[18] Later came reports of BL in European children living in Africa (Clift et al., 1963, 244; Glemser, 1971)

> We know too that it has been occurring for many years past although its unitary nature was not recognised... The exact nature of this tumour has yet to be defined and a number of names could and have been applied to it: lymphosarcoma, reticulum cell sarcoma, stem cell lymphoma. It is here referred to as "lymphoma". It is evidently a highly anaplastic malignancy developing in primitive mesenchymal cells of the reticuloendothelial system.
>
> (Burkitt and Davies, 1961, 368)

So the failure to recognise this tumour syndrome reduced its apparent incidence. This is hardly surprising. But there were more interesting manifestations of this classificatory confusion. For instance, several malignant diseases, with apparently high incidence in Africa at this time, turned out to be part of the lymphoma syndrome. These included round-cell sarcoma of both the kidney and liver, retinoblastoma, neuroblastoma and lymphosarcoma of the ovary (Higginson and Oettlé, 1960; Brew and Jackson, 1961).[19] Some cases of BL were also reported as non-malignant disease (Murray and Brandt, 1951; Bowesman, 1960).

One important aspect of this diagnostic confusion was apparent in the investigations of the relationship between BL and childhood leukaemia.

2.3.2 BL and leukaemia

Perhaps surprisingly, there was a detailed registry of cancer cases kept in the Ugandan capital at this time.[20] Of special note is the high incidence of lymphomas recorded in the area. O'Conor and Davies, 1960, reviewing the data on childhood malignancies from the Kampala cancer registry, noted the high incidence - approaching 50% of all childhood cancer - of lymphoma, and the very low rates of leukaemia. This was despite the relatively effective surveillance in place for leukaemia, an accidental feature of examination of blood-films for malaria, which would also detect the abnormal cells seen in leukaemia.

[19] It is a puzzle that, despite the distinctiveness of the syndrome, its recognition was so tricky: 'Experienced clinicians, radiologists and pathologists are of opinion that, accepting the criteria appropriate to their specialty, they can recognise this clinical entity; and it is perhaps strange that before Burkitt's observations the condition had escaped identification.' (Pulvertaft, 1964, 238). If the tumours were mistakenly classified as a variety of different conditions then the apparent incidence of each would have been artificially reduced. This diagnostic confusion is evident in many publications from this period. For example a review of head and neck tumours from Kenya in 1961 (Clifford, 1961), the author referred to forty cases of the lymphoma syndrome as instances of Hutchison's syndrome (Clifford, 1961, 710). While this term is no longer widely used, it refers to a syndrome of a primary adrenal tumour with metastases to the bones of the skull, generally caused by a neuroblastoma. It was first described by Hutchinson in 1907. Referring to Hutchison, 1907, the clinical differences between this and the Burkitt tumour are striking. Histologically, while the cells of the tumour are similar to those of Burkitt's lymphoma, both being small, round and basophilic, the overall microscopic appearance is quite different owing to the fibrillar matrix within which the cells of Hutchison's syndrome are typically embedded. Clifford recognised this difficulty: 'Histologically, the tumour appears as a lymphoma or round cell sarcoma rather than a neuroblastoma, though clinically there is nothing to suggest such a diagnosis. It is recognized that the term "round cell tumour" is inexact and unsatisfactory, but to label this tumour a lymphoma or a neuroblastoma, would be incorrect at present.' (Clifford, 1961, 716)). This was despite the very close fit in clinical appearance between the syndrome he described, and that which Burkitt described in his 1958. It is most unlikely that Clifford's cases were true neuroblastomas.

[20] For background on the Kampala Cancer Survey at the time see Davies et al., 1958. This gives very useful general information on the nature of the survey. It also gives overall rates of different types of cancers in the area of Mulago, with comparison rates from Denmark.

This high lymphoma:leukaemia ratio was the converse of that which one would expect to see in a Western population (O'Conor and Davies, 1960, 528-9). Further confirmation of the high historical incidence of lymphoma in the Mulago hospital pathology records arose as a side-effect of research into malignant diseases of the heart (Lothe and Somers, 1960).[21] They concluded, on the basis of their research and a review of the contemporary literature that: '...in the Uganda series malignant lymphoma is nearly four times as common as it is in British and American series.' (Lothe and Somers, 1960, 160).

This tallied with the observations of Burkitt and O'Conor (Burkitt and O'Conor, 1961, 259-60) that leukemias were present at a much lower incidence, and the lymphomas at a much higher incidence, than would be expected in a European population. They went on to add: 'We have not as yet seen leukemia in association with this lymphoma syndrome, even in the terminal phases of the disease.' (Burkitt and O'Conor, 1961, 268). This was further supported by an analysis of leukocytes from 32 cases of the lymphoma syndrome seen at Mulago, which demonstrated no evidence to suggest that the lymphoma syndrome was in any way leukaemic (Stansfield, 1961). Likewise, O'Conor and Davies (1960, 533) suggested that they had seen only one patient who had developed any leukaemic features in their series of 57 cases. So was the lymphoma syndrome in some way an unusual manifestation of acute lymphoblastic leukaemia? Could the two diseases share a similar aetiology, with each disease developing on the basis of prevalent local environmental factors?

> Study of the lymphomas of African children has led to the observation that these may well be simply a different expression of childhood lymphoma which in North America usually manifests itself as acute leukaemia.
>
> (Dalldorf, 1962, 1028)

Similarly, O'Conor and Davies (1960) mention the relationship of lymphoma and leukaemia in Africa:

> Since malignant lymphoma in this study has replaced lymphatic leukaemia as the most common neoplasm of childhood, it is suggested that this may represent a true variation in the site of malignant proliferation and, more important, an alteration in the natural history and course of the diseases of the lymphopoietic system.
>
> (O'Conor and Davies, 1960, 529)

Part of this environmental thinking was due to the sharp local variations seen in leukaemia and lymphoma incidences. Surveys from Africa, but outside the lymphoma belt (see section 2.3.4), such as Higginson and Oettlé (1960, 651-3), showed similar rates of leukaemia and lymphoma to those of Europe and North America. This inversion of the expected ratio of leukaemia and lymphoma cases within the belt led to the suspicion that some kind of geographically distributed environmental factor was causally responsible. When this suspicion was combined with the other unusual features of the tumour, the theory

[21] Interestingly, this work was carried out at the urging of Professor J.N.P. Davies, who features heavily in the story of the discovery of BL. (Lothe and Somers, 1960, 166).

that BL might have an infectious cause became a serious possible explanation for the aetiology of the tumour.

2.3.3 BL as an infectious disease?

It is worth at this point reviewing the evidence that supported the hypothesis that BL was an infectious disease, or one at least caused by the influence of an external environmental factor. First, the homogeneous nature of the lymphoma syndrome supported a distinct identity for the syndrome. Second, its characteristic age incidence and geographical distribution, in the absence of some identified racial or tribal (specifically behavioural) cause (Burkitt and Wright, 1963, 130-1). Third, the apparently reciprocal relationship between leukaemia and the lymphoma syndrome. Finally, it was suggested that the tumour histologically resembled certain animal tumours caused by viruses.[22] Only an infectious cause really seemed to plausibly account for these features, as the environmental factors generally considered to be plausible contributors to oncogenesis appeared not to correlate with the tumour. These factors included diet, aflatoxin exposure (see section 4.5.2.3) and background radiation levels (reviewed in Burkitt and Wright, 1963, 131ff).

This infectious hypothesis gained some further support from a previous research programme. Authors had previously noted unusual jaw swellings in this part of the world which were apparently due to the actions of an infectious agent (e.g. Bowesman, 1960, ascribed similar lesions in Ghanese children to histoplasmosis; while Murray and Brandt, 1951 outlined the possibility of a relationship between histoplasmosis and malignant lymphoma). However, as no fungi were seen in any of the material, this research had become outdated. But a viral infection could provide a very convincing explanation of the phenomena:

> Although there is no doubt that the disease as reported in this paper is a neoplastic one, and we have found no identifiable fungi in any of the material studied, the possibility of a priming action of the reticuloendothelium by some parasite with subsequent malignant change must be seriously considered.
>
> (O'Conor, 1961, 282)

Put together, this evidence suggested that the syndrome might be due to an arthropod-vectored virus (Burkitt, 1962a, 234; Burkitt, 1962d, 217-9; Burkitt, 1962f, 77). This suspicion led to the re-examination of an old direction of research: a more detailed investigation of the geography of the disease.

2.3.4 Geographical epidemiology and the tumour safari

The suspicion that BL occurred in a particular geographical pattern - soon known as the *lymphoma belt* - was an important reason for researchers to suspect an infectious aetiology for it. However, substantial refinement was required before this could be anything other than speculation. This research proceeded in two ways. First, the boundaries of the lymphoma belt were sought by attempting to discover cases of BL occurring outside the endemic area. The diagnostic confusion that existed before the clinical description

[22] See the brief review of animal viral oncogenesis in section 1.1.

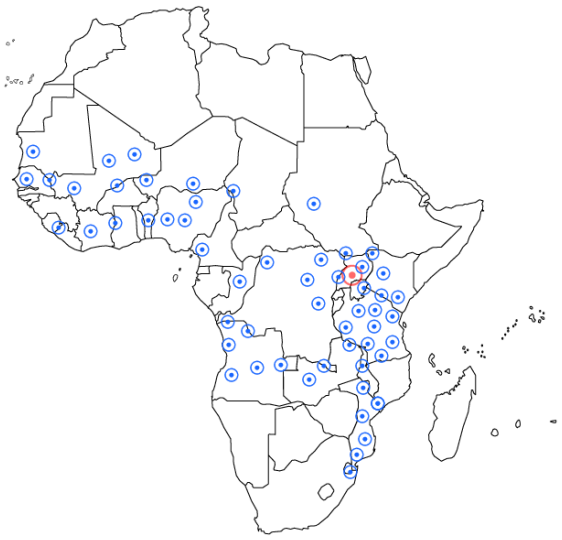

Figure 2.1: Cases of the lymphoma syndrome reported from the 1961 postal survey. Each blue dot represents a medical centre reporting one or more cases. The red dot is Mulago hospital. Data derived from Burkitt, 1962a.

of BL suggested that diagnosis could be easily be missed altogether, or fitted in to many other diagnostic categories. Thus they sought to demonstrate the scarcity of the disease outside the lymphoma belt. For instance, O'Conor (1961) stressed the relatively precise localisation of BL by reviewing the literature for cases, and finding just four plausible reports occurring outside the lymphoma belt,[23] although Burkitt later managed to increase this to fourteen.[24] All give descriptions of similar cases occurring well outside the endemic area. However, this scarcity suggested that the disease was very much less common outside tropical Africa than within it.

The second part of their research was to map out the precise spread of the lymphoma belt itself. Burkitt initially conducted a postal survey of African hospitals, asking for sightings of the distinctive clinical features of the lymphoma (Burkitt and Wright, 1963, 103-4). This was sufficiently successful to allow him to map the location of these case reports of lymphoma syndrome, as demonstrated in figure 2.1.

Thus, Burkitt (Burkitt, 1961) concluded that the lymphoma syndrome was seen to occur in a belt across the continent, spanning the equator with a tail running down the east coast and sparing the northern and southern extremities (Burkitt, 1961, 511-2).[25] He also noted that the islands of Zanzibar and Pemba, lying off the coast of Tanganyika (now incorporated into Tanzania) were spared entirely (Burkitt, 1961,

[23] These are Maxwell, 1954, Nigam, 1954, Egan and Dodd, 1957 and Orsós, 1958.

[24] These publications, listed in Burkitt and Wright, 1963, 69-70, are three cases in Dargeon, 1961, and one each in Brown and O'Keefe, 1928, Craver and Copeland, 1934, Christiansen, 1938, Salman and Darlington, 1944, Burford et al., 1944, Hellwig, 1947, Maxwell, 1954, Nigam, 1954, Egan and Dodd, 1957, Orsós, 1958 and Phan-Ngoc-Duong and Dao-Duc-Hoanh, 1961.

[25] It appeared that Burkitt had a '...lifelong passion for plotting things on maps' (Heaton, 1993, 951). His later work, on dietary fibre and its link to diseases of the developed world, also came, in part, from work on geographic pathology. This love for maps has been attributed to the influence of his father, a surveyor. His earlier work also demonstrates geographical skill, for example demonstrating geographical variations in the incidence of testicular hydrocoeles (Burkitt, 1951).

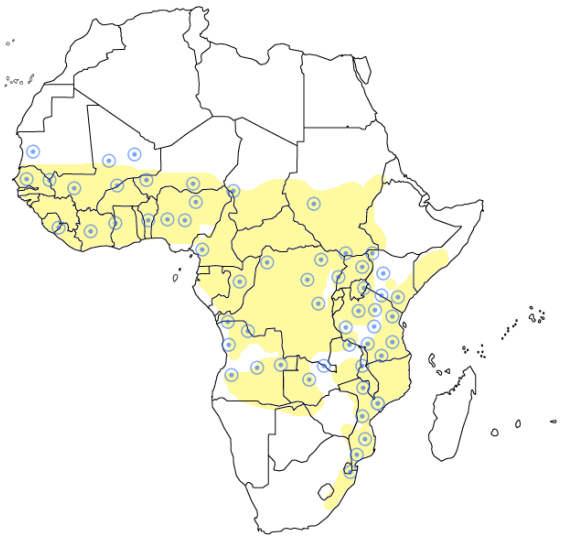

Figure 2.2: Map showing reported cases of BL (blue dots) with climatic area indicated in yellow.

512). Furthermore, there appeared to be areas within the tumour belt in which the tumour did not occur or, at least, occurred with greatly reduced frequency. This distribution led to enquiries into the unifying factors common to these areas:

> I showed my maps of tumour distribution to Dr. Haddow, of the Virus Research Institute at Entebbe, and he tried by a reverse process to see whether he could produce a map similar to mine. He found that if he took a map of Africa and eliminated all areas where the mean temperature fell below 60 degrees F. at any time of year, areas over 5,000 feet in altitude, and areas where the mean rainfall of the year fell below 30 inches, he could produce a map which is almost identical with the map of tumour distribution. This suggests that the type of vegetation (being dependent on rainfall) and the temperature are two major factors determining the distribution of this tumour.
>
> (Burkitt, 1961, 512-3)

This map of climactic factors, overlaid with Burkitt's survey cases, is given as figure 2.2. As variations in the local climate seemed to influence the incidence of the tumour so strongly, agents dependant on particular climatic conditions became implicated in the development of BL. For instance, certain arthropods were known to have distributions that appeared compatible with the lymphoma belt. Two arthropods of known medical importance seemed likely candidates (Burkitt and Wright, 1963). The first was tsetse (*Glossina* spp.), which are the insect vectors of African trypanosomiasis (Burkitt and Wright, 1963, 131-2). Distribution maps of the tsetse fly seem to fit the climate and case maps rather well, as indicated in figure 2.3.

A similar correlation was seen with *Anopheles* mosquitoes, the vector of malaria. The geographical distribution of malaria, and hence the insect, was again similar to the case distribution(see figure 2.4). A further piece of supportive evidence was also available implicating the mosquito. An epidemic of

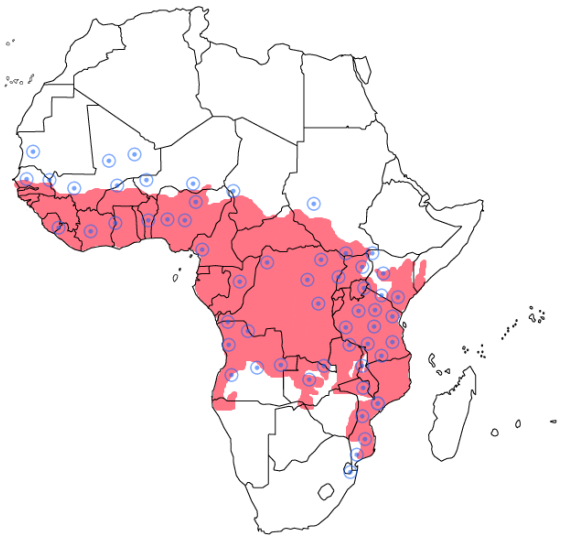

Figure 2.3: Map showing reported cases of BL (blue dots) with areas of high rates of tsetse indicated in red. Tsetse data derived from Buxton, 1955

O'Nyong-Nyong fever had occurred in Uganda at the time that Burkitt and colleagues were working. This infection appeared to be geographically limited by environmental features in a similar way to the lymphoma syndrome, and was later shown to be transmitted by *Anopheles* mosquito. Thus climatic factors had been shown, in this case, to limit the extent of a clinical disease.

Put together, the thought that the tumour syndrome was caused by some sort of arthropod-vectored agent, such as a virus, seemed highly plausible.[26] In further support of the vectored-virus theory, Burkitt and Davies argued from analogy:

> From what has been found out so far about this tumour one cannot but be impressed with the resemblance this has to some of the virus induced tumours of mammals and cannot avoid wondering if this might not be a virus induced tumour of African children and rarely of African adults. It would appear that there is a *prima facie* case for investigating this possibility.
>
> (Burkitt and Davies, 1961, 369)

This led to the *tumour safari* in late 1961 (Burkitt, 1962e; Burkitt and Wright, 1963, 113ff), where Burkitt, T. Williams and C. Nelson visited 56 medical centres across eight countries on the south-eastern tail of the lymphoma belt (as per figure 2.5).

> It was therefore decided to make an attempt to visit every medical unit along the presumed "positive" side of the tumour belt edge, and then every unit on the "negative" side.
>
> (Burkitt and Wright, 1963, 113)

[26]The implication of a vectored virus in the aetiology of the tumour syndrome was not initially Burkitt's: '…it was Professor Davies who first suggested to me that this tumour might be virus-induced and the credit for that insight goes entirely to him.' (Burkitt, 1961, 512).

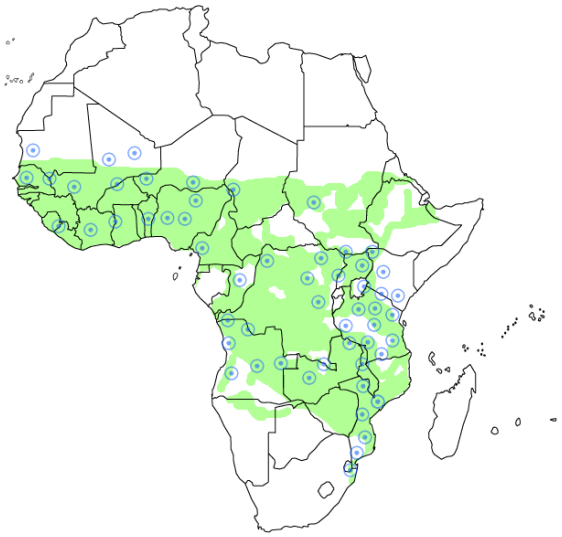

Figure 2.4: Map showing reported cases of BL (blue dots) with areas of holoendemic malaria indicated in green. Malaria data derived from MARA, 2010.

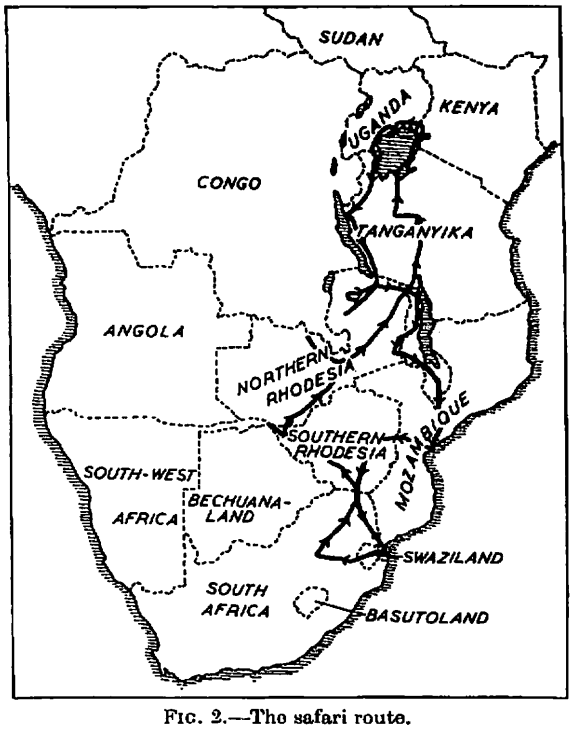

FIG. 2.—The safari route.

Figure 2.5: Route of the tumour safari. Reproduced from Burkitt, 1962e, 381

With geographical precision in mind, and given the sparse distribution of health-care in this region, the researchers attempted to improve the accuracy of the survey by recording the place of origin of the sufferer, rather than the location of the medical centre at which they were seen.

2.4 Making mechanisms: EBV

By the end of 1962 researchers strongly suspected that the tumour syndrome was caused by a virus. This weight of suspicion led to further changes in the type of investigation they conducted. Rather than investigating the possibility that the tumour could be caused by a virus, researchers instead attempted to discover which particular virus was causing the tumour, and the underlying mechanism of this causation. This phase of inquiry is therefore characterised by the construction of a specific causal mechanism for BL. Much of this section will therefore be concerned with the way that EBV was discovered from BL samples, and the way in which a causal link was made linking the two. The structure of this discovery will be familiar; again, researchers passed through the steps of suspicion, domain-finding, mechanism construction and application. But these stages attempt to answer a slightly different research question from those previously. Instead of addressing the question of what causes BL, the question under investigation was "what causes our cultured BL cells to behave as they do". This section therefore aims to demonstrate the nested nature of causal discovery in medicine, by showing that similar processes occur in component parts of causal discovery as happen in the broader case.

2.4.1 Suspicion and domain finding

As we have seen, the epidemiological features of BL, in particular its odd geographical distribution, suggested that it was caused by an infectious agent. When this suspicion was combined with contemporary discoveries of the role of viruses in causing malignancies in animals, BL became a prime target for virological investigation. The next step was to discover, if possible, which virus or sort of virus was specifically implicated. How did this come about?

2.4.1.1 The meeting at the Bland-Sutton, 1961

During his time at Mulago, Burkitt maintained informal connections with the Academic Surgery Unit at the Middlesex hospital (Epstein, 2005, 2). When he visited England in early 1961, he gave the first lectures about the lymphoma syndrome outside Africa. One of these lectures was at the Academic Surgery Unit of the Middlesex hospital on the 22nd March 1961. Entitled '*The Commonest Children's Cancer in Tropical Africa. A Hitherto Unrecognized Syndrome*' (Epstein, 2005, 2-3), it was a short review of the clinical and epidemiological features of the tumour. Anthony Epstein, a physician with research interests in tumour virology,[27] was, by chance, in the audience. He immediately became captivated, in particular by the apparent epidemiological dependency of the tumour syndrome on climatic factors. This led him to suspect that the tumour was caused by an arthropod-vectored oncovirus. This led to Epstein's immediate decision to investigate the tumour for viruses. He therefore arranged for BL material to be sent from Kampala to London by air, and began work.

[27] He began his research career working on the Rous sarcoma virus (Epstein, 1956; Epstein, 1958a; Epstein, 1958b), before moving on to work on HSV at the Middlesex hospital in London (Epstein, 1962a; Epstein, 1962b; Epstein et al., 1964c).

2.4.1.2 No virus found!

Epstein's initial investigations on the tumour material were thorough, but unpromising. Some 2 years of research passed by without the detection of a virus in the tumour cells (Epstein, 2005, 4). This was despite the use of the best available techniques, including cell cultures, infection of fertilized hens' eggs and inoculation of material into newborn mice. While no virus could be found, there were tantalising suggestions of viral activity in Burkitt's lymphoma cells when inspected using electron microscopy. In eleven tissue samples taken from two patients, Epstein reported significant evidence of viral infection,[28] but no viruses:

> The findings reported here are important therefore, since they suggest that it is unlikely that the characteristic nuclear features of the cells are connected with the presence of an hypothetical infecting virus. Certainly in the present material... there was no morphological evidence whatsoever of particles with virus-like characteristics, and this conclusion is of particular relevance for Burkitt's syndrome with its circumstantial pointers to a possible virus cause.
>
> (Epstein and Herdson, 1963, 57-8)

2.4.2 Making mechanisms

2.4.2.1 Cell Lines

Epstein remained convinced a virus was to be found. Luckily, in late 1963, he managed to obtain a substantial research grant of $45000,[29] which allowed the recruitment of Bert Achong and Yvonne Barr in the winter of 1963. With extra hands available, the decision was made to attempt to grow BL cells *in vitro*. This would presumably allow the developing cells to avoid immune system attack, permitting viral replication and hence detection. Attempting to culture the cells was an audacious move. At the time, it was considered either very difficult or perhaps impossible to grow human lymphoid cells in free suspension. Certainly, no human lymphoid cell had been maintained long-term in this manner. It had occasionally been done with other types of tumour cell. For example, researchers had established the HeLa cell line from a cervical tumour in 1951 (see footnote 19, p.71). Some types of animal lymphoid cell lines had also been successfully established. For instance, Fischer, in the late 1950s, had successfully cultured malignant mouse lymphoblasts (Fischer, 1957; Fischer, 1958). Significantly, while no virus was detectable in these cells *in vivo*, when they were cultivated *in vitro*, VLP became detectable by electron microscopy (Epstein et al., 1964d). In this case, it was thought that isolating the cells from the host immune system might permit the growth of otherwise latent viruses, and thus reveal their presence microscopically to investigators (Epstein, 2005, 4). This analogous research was an important evidential

[28] The nuclear changes that Epstein refers to were known to occur as a consequence of many types of viral infection. For instance, Leplus et al., 1961, in a review of 36 cases of Hodgkin's lymphoma, lymphosarcoma and reticulosarcoma, noted three types of characteristic nuclear changes, and suggested their similarities to those changes seen in infections with adenovirus, varicella zoster virus, ectromelia virus, encephalomyocarditis (a porcine picornavirus), molluscum contagiosum virus, Rous sarcoma virus and murine polyomavirus (Leplus et al., 1961, 2790).

[29] This was from the U.S. Public Health Service, grant number C6407 (Epstein and Barr, 1964).

prompt for the coming work. But the actual establishment of a culture line of Burkitt lymphoma cells arose by accident.

A delayed flight led to a breakthrough. The 3rd Dec 1963 BOAC Kampala flight, bringing a fresh sample of Burkitt's lymphoma tissue to London,[30], was diverted to Manchester due to bad weather. When the delayed sample was received, the usually clear transport medium had become turbid. Suspecting bacterial contamination, and hence possible spoilage of the sample, the cloudy fluid was examined under the microscope. No bacteria were apparent. Instead, the cloudiness was due to tumour cells, shaken free from the cut edges of the biopsy sample, and apparently surviving free in suspension.[31] These free cells continued to thrive, and became known as the EB1 cell line. By the time this finding was published, the cells had survived in free culture for eight weeks (Epstein and Barr, 1964, 253). This ability of BL cells to grow freely in culture was truly remarkable, as only a few human tumours possessed this property. In fact, culturability was more normally associated with animal tumours. As Epstein commented:

> Morphologically, the cells appear to be altered primitive lymphoblasts, and their unusual mode of growth without attachment to glass closely resembles that of cultured malignant mouse-cells of this type.
>
> (Epstein and Barr, 1964, 253)

2.4.2.2 Detection of virus in cultured cells

On the 24th of February 1964, samples of the EB1 line were examined using thin section electron microscopy. The findings were exciting - with '...unequivocal virus particles in a cultured BL cell in the very first grid square to be searched' (Epstein, 2005, 6). These VLPs, seen in a small proportion of EB1 cells at 75 and 82 days culture, resembled HSV, but with two important differences. First, they were significantly - about 20% - smaller, with a diameter of about 110-115nm. Second, unlike infections with HSV, they did not cause cultures to become damaged and lytic. Yet they appeared to be actively infecting the cells, rather than being a contaminant. As the first report of viral detection in the tumour cells (Epstein et al., 1964a) reported:

> ...the presence of immature and mature forms certainly indicates active virus replication. The appearances recorded are not consistent with mere virus uptake...with regard to the more important question of the relationship of the virus to the Burkitt lymphoblasts in which is has been found, it seems that a "passenger" role can be assumed since the agent has persisted *in vitro* in the dividing cells for many weeks. Any more significant connection remains to be established.

[30] This sample was taken from a maxillary lymphoma of a nine year old girl (Epstein et al., 1965a, 69).

[31] It may well have been the case that the transport medium from other specimens, too, would have contained these suspended cells. What was distinctive in this instance was simply the delay in transit, which permitted more cells to become detached from the tissue mass than usual. The particularly large number of cells in suspension led to the cloudy appearance of the transport medium, which raised the spectre of possible bacterial contamination. In order to exclude this - which would have made the tissue sample useless - Epstein examined the transport medium microscopically. This prompting to look at the transport medium under the microscope was therefore the reason that this property of BL cells was recognised at this time.

(Epstein et al., 1964a, 703)

The identification of virus by morphology alone was remarkable, and controversial (Epstein, 2005, 6).[32] The EB1 line continued to survive, and was also reported as containing detectable virus at 287 days (Epstein et al., 1964b, 1233). In summary, a virus had been detected in BL cells, which appeared to be responsible for their malignant transformation. As I will discuss in section 2.4.2.4, EBV was not the only virus to be isolated from BL cells. However, it was the only one that was regularly associated with BL tumour cells. In fact, it was detected in each of the BL cell lines that were rapidly established from diverse types of BL tissue.

2.4.2.3 More cell lines, more tumours?

At the same time as the workers in London were discovering the unusual properties of BL cells, Nigerian researchers were engaged in identical work.[33] Pulvertaft, 1964 reports the establishment of a second cell line at the University College Hospital, Ibaden, Western Nigeria: 'At the time of writing one strain has been maintained in subculture for ten weeks, and a second for five weeks, without change of characteristics.' (Pulvertaft, 1964, 239). These cells, which later became known as the Raji line, had similar cytological features and characteristics to the EB1 line, and contained similar viruses on electron microscopy, as described in Epstein's report (Epstein et al., 1966).

Following the reports of the establishment of the first two cell lines - EB1 in London, and Raji in Ibaden - and the reports of an unknown virus in them, subsequent cell lines were established. A further two were made by the Epstein lab - EB2 (Epstein et al., 1964b) and EB3 (Epstein et al., 1965a), while a fifth cell line - SL - was reported in Stewart et al., 1965. See table 2.1 for an overview of these early cell lines.

The first notice of the EB2 line was Epstein et al., 1964b. This detailed the detection of VLPs in a second cell line, originating in tissue taken from an ovarian tumour from a 7 year old with abdominal and left-sided facial tumours. Again, 1-2% of cells contained a morphologically similar virus to the maxillary-origin EB1 line. The cells themselves had virtually identical characteristics to the EB1 line. The authors thus concluded that culturability was a general property of BL cells. As with EB1, they did not (yet) ascribe a causal role to the virus in immortalization, but:

> ... the failure of extensive attempts to isolate and identify this agent clearly indicates that it is not merely some banal "passenger" virus such as herpes simplex which it somewhat resembles morphologically.
>
> (Epstein et al., 1964b, 1234).

Furthermore, this virus appeared to share similarities with other known viral pathogens beyond its gross appearance:

[32] This was, incidentally, the first case in which a virus was discovered solely by electron microscopy (Epstein, 2005, 7). It is worth noting Epstein's familiarity with using the gross structure of viruses for classificatory purposes (Epstein, 1962a).

[33] In fact, the two teams reported their findings simultaneously in the February 1st 1964 issue of the Lancet

Line name	Date	Patient details	Tumour location	Geographical origin	Publication
EB1	Dec 1963	9 yrs, female	Maxillary	Uganda	Epstein and Barr, 1964
Raji	Late 1963	Various	Various	Nigeria	Pulvertaft, 1964
EB2	Sep 1964	7 yrs, female	Ovarian	Uganda	Epstein et al., 1964b
EB3	Late 1964	3 yrs, male	Temporal	Uganda	Epstein et al., 1965a
SL	Dec 1964	7 yrs, male	Mandibular	Nigeria	Stewart et al., 1965

Table 2.1: Summary of early BL cell lines, showing original designations, characteristics of originating material, and first report.

> ...the virus found...is in many respects similar to that in the Lucké frog neoplasm[34], and in addition recent fluorescent antibody studies have shown that these virus-carrying lymphoblasts have very close antigenic affinities with a murine neoplastic virus and a virus associated with human leukaemia.
>
> (Epstein et al., 1964b, 1234)

This fluorescence work is worth briefly reporting. Fink et al. (1964) demonstrated that antigen from the as-yet unidentified virus, when tested against cells from other malignant diseases, showed reactivity against both blood cells and bone marrow from patients with leukaemia. Further, antigen prepared from Rauscher murine leukaemia[35] reacted against BL cells in culture. This led to the suspicion that EBV caused other tumours in humans.

The 'extensive attempts' (Epstein et al., 1964b, 1234) made to isolate the virus by Epstein, Achong and Barr is no hyperbole. As they later recounted in the paper detailing EB3, they attempted the following using EB1 and EB2 material:

> ...various kinds of virus preparation from both strains of lymphoblasts were inoculated on a large scale into groups of developing chick embryos, litters of newborn hamsters, litters of newborn C57 brown mice, litters of newborn Swiss mice, and into tissue culture systems: nine different test culture systems have been tried for every virus preparation and have each been carried through eight blind passages. All these biological tests have been uniformly negative, as have tests for mycoplasma, and the virus would therefore seem to be a new unusual unidentified agent of unknown nature.
>
> (Epstein et al., 1965a, 70-1)

This publication then goes on to describes the new EB3 line. EB3 originated from a tissue sample taken from a 3 year old boy with temporal and maxillary tumours (Burkitt case J258, Epstein et al., 1965a, 74). After about two months, a morphologically, kinetically and ultrastructurally identical strain to EB1 developed, although the rate of detection of VLPs was slightly higher, with virus found in roughly 3-4% of the cells.

[34] See section 1.1

[35] See section 1.1

The SL line originated from a tissue sample from a 7 year old Nigerian boy (D.A.), who had been treated for BL at the National Institute for Health in Bethesda, Maryland. They noted that the virus seen was similar to Epstein's, suspected that it was probably a type of herpesvirus, but excluded the possibility of it being HSV, on the grounds that it displayed no cytolytic effect (Stewart et al., 1965, 320). They also noted that it appeared to cause a distinctive cytogenetic alteration in the cultured cells (Stewart et al., 1965, 321; Jacobs et al., 1963).

What did this work on cell lines do? First, it demonstrated that BL was unusual in the respect that cells of the tumour could survive long-term in free culture. This property appeared to be a general one of the tumour. We see this in the observation that various tissue samples taken from different anatomical and geographical sites were similarly capable of cultivation *in vitro*. This also provided evidence that the clinical syndrome really was a unified one, probably arising from a common cause. Second, it demonstrated similar VLPs in each of these tumour lines. This in turn suggested that the viruses seen in the cell lines were not mere contaminants (say, oral flora with a propensity to live in tumour cells). Third, it suggested that the virus was, in some way, aetiologically responsible for the transformation that made it possible to culture BL cells. This was, in part, dependant on the analogical research conducted on non-human animal tumour cells. Fourth, it sparked the investigations of the possible role of this unknown virus in other diseases. All this fruitful investigation occurred in the absence of a clear, distinct description of the virus. This was soon to follow.

2.4.2.4 Describing the virus

Epstein et al., 1965b has been cited as the first real description of EBV (Miller, 1974). First, they review and discuss other possible causes of BL. It is worth digressing slightly here in order to mention these other agents:

Reovirus 3. An agent closely related to reovirus 3 (CAN 230) was first detected by serial cell cultures in a single Ugandan BL sample (Bell et al., 1964). Further, a small serological survey revealed antibodies to this virus in almost all individuals with BL tested. This was extended to ten related reovirus strains in further work (Bell et al., 1966). As with EBV, the original BL tissue was taken from a variety of anatomical and geographical sites. Sero-prevalence surveys, for antibodies against three reovirus strains, were undertaken in areas of Africa with differing BL incidence. This also included testing the sera of fourteen children with the disease. This work showed generally low incidence of antibodies against reovirus 3, except in the BL group (Levy et al., 1968). A possible synergy was also suspected between reovirus 3 and HSV. Reoviruses were also known to cause murine lymphomas (Stanley et al., 1966; Joske et al., 1966). These murine lymphomas also shared cell-surface markers with BL (Stanley and Keast, 1967). A very similar virus was also known to cause plant tumours (Streissle and Maramorosch, 1963). Reovirus was a plausible cause for a number of other reasons, not least because it was also known to be carried by several species of mosquito (Parker et al., 1965).

Arbovirus. It was suggested very early in research into BL that an arbovirus[36] could be responsible. For example, Haddow and McCallum (1962, 36) noted that the epidemiological character of the

[36] Arboviruses are viruses that are transmitted by arthropods

tumour suggested that this agent was at least a plausible one '...up to the time of writing, nothing has been found to preclude the possibility that Burkitt's syndrome could be caused by an arthropod-borne virus.' A summary of the evidence was reviewed by Haddow (1964). He suggested, amongst other factors, that BL must be an uncommon outcome of a common infection. This, in combination with the geographical evidence, suggested a mosquito-borne viral infection. In concert with the characteristic age-incidence of the tumour syndrome, this evidence led him to suggest an analogy with yellow fever, a common, mosquito-borne arbovirus. So arbovirus was initially suggested to cause BL on the grounds of plausibility and analogy. While the view that arboviruses alone were the cause of BL fell out of favour, the suspicion that they are somehow causally implicated in BL pathogenesis has recently been somewhat resurgent. In very brief summary, arboviruses have been shown to be potentially oncogenic, occur in an epidemic fashion and apparently temporally precede BL case clusters (van den Bosch, 2004, 742). More, the likely age at first arbovirus infection is a good match for BL (van den Bosch, 2004, 741). It is therefore suspected that arboviruses can act as potential co-factors for BL oncogenesis (van den Bosch, 2004, 742).

Echo 11. This virus was detected in two cases of BL, as well as in three other types of tumour (Munube and Bell, 1967). However, no further research program seems to have come about based on this finding.

HSV. Woodall and Haddow (1962, 30) reported the isolation of a herpes virus of unknown type from BL biopsy material. Regrettably, this paper is often cited as a report of - specifically - HSV isolation. Other reports (e.g. Simons and Ross, 1965; Woodall et al., 1965) are similarly equivocal. Epstein posited contamination from the nasal cavity, for instance, as a possible source for these unknown herpesviruses (Epstein et al., 1965b).

Other agents. These included *Mycoplasma* (Dalldorf et al., 1966), and multiple other filterable agents (Dalldorf and Bergamini, 1964; Dalldorf et al., 1966; Munube and Bell, 1967)

The evidence supporting the role of EBV in BL, though, was different from that supporting these other agents. EBV was not merely found in tissue samples (although one group did manage to directly detect it, and a range of other candidate viruses, in BL biopsy material (Griffin et al., 1966)). Instead, it could grow in tissue culture systems (Epstein et al., 1965b, 766). The many instances in which this virus was detected prompted a morphological description. Two forms of VLP were seen. The first was a 75nm diameter, hexagonal, enveloped particle, with a single membrane and an empty or a ring-shaped inner body. This was found in large quantities in both the nucleus and cytoplasm of affected cells. The authors suggest that it represented an immature form of the virus. Secondly, and more rarely found, were larger 110-5nm diameter particles, which were bienveloped and contained a dense nucleoid of diameter 45-50nm. These were invariably cytoplasmic (Epstein et al., 1965b, 763). In addition, there was also some evidence for the budding of particles as part of maturation, in a similar fashion to HSV (Epstein et al., 1965b, 764-5). These were found exclusively in dead or abnormal cells, and in strict association with cellular structural changes. Epstein et al. (1965b) claimed that the virus was both novel and distinct from HSV. Yet they appeared wary of making aetiological causal claims for this new virus:

> The fact that this unknown virus has now been observed in each of three lines of cultured Burkitt tumour lymphoblasts examined by electron microscopy is perhaps no more than a coincidence; this type of cell could be picking up as a passenger a particular unknown non-pathogenic agent harboured by patients with BL. Alternatively, the cultured cells may have grown *in vitro* just because they are persistently infected with this agent, which might be responsible for their growth potential.
>
> (Epstein et al., 1965b, 766)

2.4.3 Application: The Henles, transformation, seroepidemiology and infectious mononucleosis

The virus found in the various cell lines remained apparently inert (Epstein, 2005, 7). While it was morphologically similar to other herpesviruses, various attempts at propagation or transmission *in vivo* were unsuccessful. This failure led Epstein, who had previously collaborated with Gertrude and Werner Henle in Philadephia, to invite them to investigate the virus. Their initial *in vivo* research suffered similar difficulties to Epstein's (Henle and Henle, 1966b, 1248). However, their *in vitro* investigations were much more successful. They first demonstrated both the immunological (Henle and Henle, 1966a; Henle and Henle, 1966b) and biochemical (zur Hausen et al., 1970) uniqueness of the virus and named it (Henle et al., 1968). Immunofluorescence testing revealed that viral antigens were immunologically reactive against the serum of patients with BL, and against BL cells. As might be expected, the extent of immunofluorescence correlated with virus particles as seen using electron microscopy (Henle and Henle, 1966b, 1252). Mystifyingly, though, despite the relatively specific immunological response obtained,[37] it was also immunologically reactive against the sera of non-BL individuals, including leukaemia patients (Henle and Henle, 1966b, 1249), individuals with nasopharyngeal carcinomas (Henle et al., 1970) and, most confusingly of all, against many normal North American control individuals (Henle and Henle, 1966b, 1250-1).[38] This suggested that infection with EBV was far more common in non-African populations than previously expected. It also suggested that, far from being specific to BL, EBV infection might cause a range of other diseases, or might cause no clinical disease at all.

However, support for an aetiological role for the virus in the development of cancer came from successful cellular transformation experiments. When BL cells containing virus were killed by irradiation, and inoculated into normal human white blood cells, these cells became transformed into an immortalized cell line (Henle et al., 1967). However, this did not happen with BL cells that were microscopically free of the virus. This experiment, in many ways complementary to the original establishment of cell lines from BL tissue, was also a methodologically significant step, as it allowed researchers to cultivate EBV outside BL cell-based culture systems. It also suggested a possible mechanism for BL development, which gained further support by the recognition of chromosomal abnormalities in the cultured cells (Henle et al., 1967, 1065), similar to those found in BL (Miles and O'Neill, 1967).

[37] There was no staining against most other suspected candidate agents, including reoviruses, with the partial exception of HSV, which had equivocal results.

[38] About 30% of children and 90% of adult controls reacted.

The high proportion of apparently healthy individuals displaying evidence of EBV exposure required explanation, however, if the aetiological role of EBV in BL was to be confirmed. Various work suggested that the virus had a global distribution, with a similar age-at-infection profile to other childhood disease. However, no likely clinical manifestations of childhood infection with EBV was known, and it required a fortunate coincidence for one to be identified. While conducting a series of surveys for EBV antibodies using sera from a variety of paediatric illnesses of unknown, but presumptively viral, aetiology, one researcher by chance developed IM.[39] It was discovered that, following this infection, her leukocytes became capable of growing in cell culture, a property they had not possessed when tested before the development of the disease. While she initially did not display antibodies to EBV, from the outset her cultured cells expressed EBV antigens, and contained the chromosomal abnormalities characteristic of BL (Henle et al., 1968, 95-6). When further sera from IM patients were examined, they were all found to be strongly positive for anti-EBV antibodies. This led to the conclusion that IM was causally related to EBV:

> Patients with infectious mononucleosis regularly develop antibodies to the herpes-type virus (EBV) found in cultures derived from Burkitt's tumors or other cells of the hematopoietic system... The epidemiology of IM and the seroepidemiology of EBV share many features. Thus, it appears that EBV, or a close relative of it, is the cause of IM. This conclusion does not preclude the possibility that EBV might also be involved, either directly or indirectly, in the etiology of Burkitt's lymphoma.
>
> (Henle et al., 1968, 101)

This probable aetiologic relationship could therefore explain the serological evidence of past EBV infection in otherwise healthy individuals. However, the causal mechanism linking BL and EBV was still rather incomplete. For one, it seemed that intact EBV particles could not be detected in biopsied (rather than cultured) BL cells. However, EBV genetic material could be found when these cells were examined using DNA-DNA hybridisation. When EBV genomes were purified and used to probe BL cells, annealing took place, indicating the presence of EBV genetic material within the host cell genome (zur Hausen and Schulte-Holthausen, 1970; zur Hausen et al., 1970). This genome integration suggested a causal, rather than passenger, role for EBV.

In summary, many kinds of laboratory and epidemiological evidence supported the existence of a causal role for EBV in BL. First, the common clinical, pathological and histopathological features of BL suggested a unified cause for the disease. These features also indicated that the disease was a malignant process. However, epidemiological findings suggested the possibility that it was an infectious disease. Next, the unusual ability of BL cells to grow in continuous culture, and the subsequent recovery of a virus from many of them, implicated a virus as the cause. This offered researchers the possibility of reconciling the apparently contradictory epidemiological and pathological features of the disease. This virus, when used to infect other lymphoid cells, caused cellular transformation, pointing to the likely role of the virus

[39] She was Elaine Hutkin, and received the thanks of the Henles in an acknowledgement to: '... patient E. H., who provided the main clue' (Henle et al., 1968, 101)

in causing malignant transformation *in vivo*. Other confusing epidemiological features - in particular the high prevalence of anti-EBV antibodies compared to the very low incidence of BL - were accommodated by the immunological and serological findings implicating EBV in diseases other than BL. Finally, EBV was shown to become incorporated into the host-cell genome, and to cause chromosomal abnormalities. This suggested the possible nature of the detailed causal mechanism linking infection with EBV and the development of malignancy *in vivo*.

2.5 Application: prospective epidemiology

The last part of the early history of EBV and BL involves a return to epidemiological research. As the summary above suggests, the causal mechanism for the role of EBV in BL was still incomplete by the early 1970s. There was, therefore, much resistance to these causal arguments at this time.[40] Various pieces of epidemiological evidence were acquired that indicated that populations at high risk of BL tended to have higher levels of anti-EBV antibodies, and tended to acquire them younger than individuals in populations at low risk (see, e.g., Henle et al., 1969). In fact, it was not until the completion of a large prospective seroepidemiological study in the lymphoma belt that it became generally accepted. This study (de-Thé et al., 1978) began in 1971 and followed 42 000 children from birth to 8 years old in the West Nile District of Uganda. Serum samples were taken at enrolment. This group was then observed. In total, 14 of the 42 000 participants developed BL at some point during the trial. These individuals had higher levels of anti-EBV antibodies than controls at baseline (de-Thé et al., 1978, 759), suggesting that EBV infection preceded the development of BL by some time.

This specific increase in antibodies preceding the development of BL, when combined with the other evidential features of the disease and of the virus discussed above, was generally seen as a conclusive demonstration of causation of BL by EBV. To conclude our survey of evidence:

> The virus can be isolated from tumour cells, virus-associated membrane antigen can be detected on the surface of tumour cells (Klein et al., 1966) and viral DNA is present in tumour derived cell lines (zur Hansen et al., 1970). All children with BL in endemic areas have antibodies to viral antigen and the geometric mean antibody titres are higher in patients than in controls (Henle et al., 1969). Moreover, prospective studies on 45,000 children in the West Nile region of Uganda showed that the 14 children who developed the tumour had acquired antibody to the virus from seven to 54 months before the lesion was detectable and that the VCA (viral capsid antigen) titre in these children exceeded the titres found in nearly all the controls (de Thé et al., 1978).

[40] As Epstein noted of this change in attitudes towards a causal role for EBV in malignancy: 'This change in perception has come about slowly and painfully; once EBV was accepted as a virus it was only with the later advent of the unequivocally oncogenic herpesviruses of Marek's disease... and New World non-human primates... that its possible function as a human tumour virus received its first tentative consideration. Even after the accumulation of abundant data on the biology and powerful lymphocyte transforming ability of EBV ..., and the persuasive evidence from the massive World Health Organization (WHO) prospective 7-year study of 42 000 Ugandan children, the reluctance to believe there could be a carcinogenic virus of man persisted.' (Epstein, 2001b, 111)

(Hutt, 1981, 762)

Chapter 3

Cervical Cancer

3.1 Introduction

This story of causal discovery should be seen as complementary to that of BL and EBV. Why then include both? First, while both tumours are caused by viruses, they are generally dissimilar. Cervical cancer, unlike BL, is a fairly typical tumour: it occurs relatively commonly, has a worldwide distribution and (generally) increases in incidence with age. Before about 1940, cervical cancer was believed to share a similar degenerative-chronic aetiology with most other malignant diseases.

This is the first difference between the cases. The second is concerned with differences in the manner of discovery of the causal mechanisms for each tumour. With BL, as we have seen, a large part of the story of causal discovery was the process of clinical discovery in the first instance. Thus, I focused on telling the parts of the causal story dealing with the suspicion of mechanism, and the process of domain finding. Here, these two phases will play a minor role. Instead, I will concentrate on the linked processes of mechanism construction and application. I will attempt to underline the differences between two different attempts at drawing a causal mechanism that occurred in the history of cervical cancer.

So this story begins with an understanding of cervical cancer as a chronic, degenerative disease. However, research on disease populations between about 1940 and 1960 suggested that this was not, in fact, the case. Cervical cancer actually had extremely unusual risk factors for a malignant disease. A great deal of sound epidemiologic data appeared to suggest that cancer of the cervix behaved like a sexually transmitted disease, and a complex web of socio-economic factors, particularly those indicating social class, marital and sexual habits and features of male sexual partners, appeared to modify the risk of developing the disease. Thus the story of the identification of these risk factors will form the first part of this chapter.

As a result of the recognition of these risk factors, and of the recently identified role played by viruses in the causation of several tumours, including BL, a virus was thus seen as a highly likely cause of the disease. Again, largely by analogy with these other viral causes of cancer, infection with HSV was suggested as the specific cause of cervical cancer. A great deal of mechanistic and statistical evidence was generated in support of this hypothesis, and between about 1970 and 1985 HSV was commonly accepted as the cause of cervical cancer. The story of the causal identification of HSV will thus form the

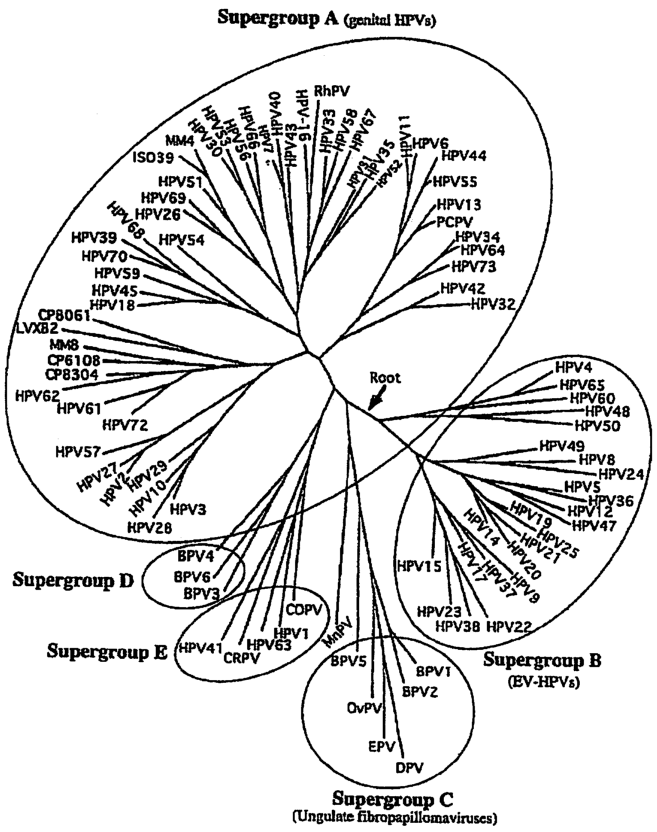

Figure 3.1: HPV phylogenetic tree (Lowy and Howley, 2001, 2232). Here, the distance from the root position is determined by the degree of sequence variation of a short section of the L1 open reading frame.

second part of this chapter.

This causal story has a twist. While we still believe that cervical cancer is caused by a virus, we now think that this specific virus is human papillomavirus (HPV), rather than HSV. In fact the causal association is so strong that HPV has been proposed as a necessary cause of cervical cancer (Bosch et al., 2002, 244). This change in causal thinking will form the final part of this chapter.[1]

HPV is a virus of the genus *Papillomavirus*, family *Papovaviridae*. It has five super-groups, categorised by sequence homology as super-groups A-E (McCance, 2004, 661), as shown in figure 3.1. Super-group A infects humans, and includes the types of the virus responsible for cervical cancer and genital warts. Viruses of supergroup B cause the rare skin disease epidermodysplasia verruciformis (EV). This disease played an important role in early cervical cancer research, as discussed in section 3.6.1.2.

[1] I will here give enough detail of the current state of knowledge of cervical cancer to sustain the historical discussion. For further information about cervical cancer epidemiology see Parkin et al., 1984; Parkin et al., 1993 and IARC Working Group on the Evaluation of Cancer Preventive Strategies, 2005. For the current position on screening for the disease, see Andrae et al., 2008; Savage, 2008. For a review of the role of papillomaviruses in cervical cancer causation, see zur Hausen and de Villiers, 1994; Bosch et al., 2002, 250. McCance, 2004 is an detailed review of HPV virology, while zur Hausen, 1989b; Howley, 1991; IARC Working Group on the Evaluation of Carcinogenic Risks to Humans, 1995; McCance, 1998; IARC Working Group on the Evaluation of Carcinogenic Risks to Humans, 2006 and Paavonen, 2007 are excellent reviews of the current mechanism of causation of cervical cancer by HPV.

Supergroup E causes cutaneous warts (benign skin papilloma), which were also vital to this causal story, as I discuss in section 3.6.1.1. Groups C and D do not infect humans, but are responsible for a variety of animal diseases, some examples of which I will mention in section 3.6.1.3. Each type of the virus is biologically distinct. For instance, they are relatively specific in terms of the species they affect, the anatomical sites which they prefer and the types of disease they cause. They are also, largely, immunologically distinct. Acquired immunity to one type does not, in general, provide immunity to others. This complex organisation of types becomes extremely important later in the story, as discussed in section 3.6.2.

The viral genome is double-stranded DNA virus of 7900 base pairs in a supercoiled circular configuration. The viral capsid is 52-5nm in diameter, consisting of an icosohedral arrangement of 72 capsomeres. Other than structural proteins, the virus encodes a number of 'early' or E proteins, three of which, E5, E6 and E7, appear causally important in oncogenesis. Together, expression of these proteins causes disruption of the normal cell cycle, possibly via interactions with tumour protein 53 (p53) and the retinoblastoma gene product (pRb). I give a brief review of the mechanism by which HPV is thought to act in section 3.6.3.

3.1.1 Cervical cancer as a chronic, non-infectious disease

Unlike the previous chapter, I will be extremely schematic when describing the suspicion that cervical cancer was a unified clinical entity, as the process of syndrome recognition is less interesting in this case. In large part, I suggest this is due to the substantial clinical cohort of patients which would have been apparent to clinicians and researchers, something quite unlike the situation with BL. Just for example, data from the USA at the beginning of the period of interest shows about 20,000 new cervical cancer diagnoses, leading to about 8,500 deaths, per year. These figures reflect a total population of about 120,000 women with invasive disease and a further 250,000 with carcinoma *in situ* (all Stamler et al., 1967, 802).

Cervical cancer was initially considered to be a type of uterine cancer, but became a clinically distinct entity in the late nineteenth and early twentieth century (Mengert et al., 1949). Thus in the following sections, while cervical and other uterine cancers are usually clearly distinguished, there may be occasional instances of classificatory confusion. For instance, publications from the middle decades of the twentieth century based upon historical data collected much earlier, particularly that dealing with retrospective surveys of epidemiological data, may be unreliable. For example, Kennaway, 1948, a review of aetiological factors responsible for cervical cancer, features some data where cervical and uterine body cancers are distinguished, and other data where the two are conflated together into cancer of the uterus. However, because of this early classificatory division, I suggest that this minor point did not present a significant issue in this causal story.

What is of interest is the nature of the causal processes suggested. Although the disease was well-known, it was generally thought before about 1960 that it was caused in a similar way to most malignant disease. For instance:

> Long-continued irritation and chronic inflammation are probably the conditions which

pave the way for the development of the new growth.

(Deaver and Reimann, 1931, 383)

Similarly, cervical trauma during childbirth was identified as a common cause of cellular irritation (Anonymous, 1919), implicating childbirth as a cause of cervical cancer. The suggestion that childbirth increased the risk of developing cervical cancer would persist. However, the future direction of cervical cancer epidemiology would not concentrate on investigating the nature of cellular irritation. Instead, it would focus on the act of childbirth itself, and its related activities, as possible causes of the disease.

3.2 Suspicion: Cervical cancer as a venereal disease

By the 1960s it was becoming apparent that, instead of behaving like most tumours, the risk factors for cervical cancer resembled those of a sexually transmitted disease. One important driver of this realisation was the introduction of cervical screening programmes in many parts of the developed world. For instance, one of the major reviews of the epidemiology and aetiology of cervical cancer at this time (Aitken-Swan and Baird, 1966a; Aitken-Swan and Baird, 1966b) derived from the screening programme in Aberdeenshire introduced in 1960. As a blunt summary of this aetiological work, the authors comment:

> The cancer patient is characterised by more marital misadventures, divorce and separation, more pre-marital coitus and deliveries and more sexual partners.
>
> (Aitken-Swan and Baird, 1966a, 656)

I will therefore briefly outline these risk factors, and the evidence supporting them, below.[2]

3.2.1 Sex

This arose from two linked observations. First, that a preventive factor seemed to operate in groups of women who did not engage in sexual intercourse. In fact, those who practice life-long abstinence are essentially at no risk of the disease. The very low incidence of cervical cancer in nuns has been known since at least the middle of the nineteenth century. For instance, Rigoni-Stern working in 1842 (available in English translation as Stavola, 1987) described a long historical series of cases in Verona (1760-1839) in which much higher rates of cervical cancer occurred in married women than in nuns. This finding was later repeated (Gagnon, 1950; Towne, 1955 and Fraumeni et al., 1969). Further work demonstrated the low incidence of the disease in monogamous populations (Gardner and Lyon, 1977; Sumithran, 1977).

Second, women with multiple sexual partners had an increased risk of developing the disease compared to controls (Wynder et al., 1954; Stocks, 1955; Jones et al., 1958; Stern and Dixon, 1961; Terris et al., 1967). Two groups were important; first, commercial sex workers, second, divorced and remarried women (Martin, 1967, 804). Both had higher rates of the disease than expected.

[2]Much of the material on cervical risk factors from this period is repeated between publications. An exceptionally useful general summary is Boyd and Doll, 1964 - which presents a case-controlled study of these risk factors in Britain. Their findings are, in general, the same as the following analysis.

As a consequence of these observations of related risk- and protective- factors, by the mid-1960s, it was thought that some covariate of sex, or of sex with multiple partners, was causally responsible:

> ... mere frequency of intercourse is unlikely to account for the differences found. Intercourse must be further analysed for special features which conduce to cancer.
>
> (Elliott, 1964, 232)

As we will discuss, the nature of other cervical cancer risk factors support this view. However, no causal mechanism or candidate causal entity was immediately suggested as a consequence of the discovery of the link with sex.

3.2.2 Marriage

Marriage increases the risk of cervical cancer. By 1950 any direct causal effect of marriage itself was generally seen as implausible.[3] Instead, married women tended to have higher rates of the other identified risk factors than unmarried ones, including having more sexual partners and more children. This was particularly the case in women who remarried after divorce. While their increase in cervical cancer risk could therefore be explicable in terms of marriage being covariate with other risk factors, the picture was slightly confused by the apparently reciprocal relationship identified between cervical and other uterine cancers. Married women had higher rates of cervical cancers than unmarried ones, but lower incidence of cancers of the uterine body (Aitken-Swan and Baird, 1966b, 628-9).

3.2.3 Parity

High parity (number of births) increased cervical cancer risk (Aitken-Swan and Baird, 1966b, 633ff). It was unclear how this related to other cervical cancer risk factors as parity seemed to bear a highly variable relationship to youth at marriage, faith and so on.

3.2.4 Faith

Disease incidence appeared to be modified by membership of certain faith groups. For instance, in a series of epidemiological surveys across Europe, North America and Palestine, members of the Jewish faith had attack rates roughly half that expected in European Christian women. Similar reductions in risk were seen in Moslem women compared to Hindu women in studies on the Indian subcontinent (Kennaway, 1948, 177-193). This is despite the high parity and generally low socio-economic status of both of these groups (Kennaway, 1948, 202-3).

Four main explanations for this reduction in risk were offered. First, perhaps the genetic origin of these groups reduced their predisposition towards developing the disease (Kennaway, 1948, 183; 204). Alternatively, perhaps this risk reduction was due to the observance of religious rituals around sexual intercourse. For instance, the Talmudic proscription of sexual intercourse (Niddah) in the first half of the menstrual cycle, from the onset of menstruation to approximately two days before the expected date of ovulation (Kennaway, 1948, 197-9; 208), and similar ritual ablutions required by Moslem law

[3] 'No one would consider the mere ceremony of marriage to have a bearing on the causation of the disease' (Lombard and Potter, 1950, 1330)

(Kennaway, 1948, 200), might have been causally responsible for the reduction in risk. Thirdly, perhaps the tendency of these groups to be highly monogamous and to practice endogamous marriage reduced the chance of contracting a venereal agent (Martin, 1967, 813). Finally, perhaps this reduction in risk was due to the ritual circumcision practised by both these groups, as I discuss in 3.2.10.

3.2.5 Youth

Relative youth at marriage, first pregnancy or first sexual intercourse increased the risk of eventually developing cervical cancer (Kennaway, 1948, 203-4; Stamler et al., 1967, 804; Meisels et al., 1977, 3078). Research into this risk factor was complicated both by its relation with other risk factors, and by interdependence between the youth risk factors. Women who had their first sexual experience when unusually young tended to both marry and have children unusually young (Aitken-Swan and Baird, 1966b, 633ff). But they also tended to have more sexual partners, to have more children, to suffer other sexually transmitted diseases and to belong to lower socio-economic classes.

3.2.6 Syphilis

The link between cervical cancer and syphilis had been known for some time. Several authors noted that syphilis appeared to modify the natural history of cervical cancer - women with both syphilis and cervical cancer tended to develop malignancy younger, more severely and with a worse prognosis (Harding, 1942; Morris and Meigs, 1950; Røjel, 1953; Wynder et al., 1954; Elliott, 1964, 233).

In all these cases the precise nature of the association is unclear: authors variously question the specificity of the Wassermann test for syphilis (Harding, 1942)[4], suggest a directly causal role for syphilis in the development of cervical cancer or suggest that syphilis may simply be an exposure marker to another agent with a causal role in cervical cancer. Familiarly, this research was complicated by the equivocal relationship of syphilis to other known risk factors.

3.2.7 Class

Women from lower socioeconomic classes had higher rates of the disease than wealthier women (Kennaway, 1948, 193-7; Elliott, 1964, 232; Aitken-Swan and Baird, 1966b, 631; Stamler et al., 1967, 802). As with the other strands of this investigation, the research was heavily confounded by the unknown relationship that class bore to other identified risk factors. For instance, women of lower class tended to have sex and marry younger, have more sexual partners, more children, more venereal disease and tended to belong to particular faith groups. Other factors were also suggested as part of the class phenomena. The disease seemed to occur at higher rates in cities than in the countryside (Aitken-Swan and Baird, 1966b, 627). There was also much variability within social groupings depending on the woman's partner's occupation (Aitken-Swan and Baird, 1966b, 639).

3.2.8 Race

Based on a study in the USA, African-American women were found to have significantly higher rates of invasive and *in situ* disease than their white counterparts (Stamler et al., 1967, 801-2). They also appeared to have a significantly worse prognosis once the disease was established. While incidence rates

[4] i.e. that cervical cancer in some way caused a false positive when testing for syphilis

were 50-100% higher, the mortality rates for African-American women were more than double those expected for white women (Stamler et al., 1967, 796-9).

3.2.9 Age

Increasing age appeared to be, in general, a risk factor for the development of the disease (Stamler et al., 1967, 799). This finding was, however, complicated by the apparent bimodal incidence of cancer of the cervix, with incidence peaks at about 40 and about 60 years. Two explanatory strategies were offered. First, these two peaks actually represented aetiologically distinct conditions with similar pathologies (Kennaway, 1948; Elliott, 1964). Alternatively, this bimodality may have reflected a degree of variability in the disease course (Kashgarian and Dunn, 1970). While interesting, I will ignore this complication.

3.2.10 Male factors

As work continued on these risk factors and their relationships, the role of male factors in the development of cervical cancer became increasingly important. This 'male factor' - that women were at increased risk of developing the disease if their husband was either promiscuous or had been previously married to a women who developed cervical cancer - was unpicked in a variety of ways.

For example, the wives of circumcised men tended to have lower incidence of cervical cancer (Kennaway, 1948, 204-5). This reduction did not seem to depend upon the reason for which circumcision was performed, as wives of men circumcised for either ritual and medicinal reasons appeared at similar risk. Thus circumcision, and its relationship to faith, was an important preventive factor. One important supportive finding was the preventive role of barrier contraception (Kennaway, 1947; Kennaway, 1948; Wynder et al., 1954; Wynder, 1955; Khanolkar, 1958; Elliott, 1964; Aitken-Swan and Baird, 1965).

It was suggested that this risk reduction was predicated upon the reduction of smegma that circumcision produced. Smegma, a deposit of secretions from the penis, was mechanistically investigated as a potential cause of cervical cancer in experimental animal systems (Plaut and Kohn-Speyer, 1947; Pratt-Thomas et al., 1955).

Considered too were notions of the partners degree of 'cleanliness' ('...cleanliness of the male partner seems to be the most important single factor' (Kennaway, 1948, 234)). While it is difficult to interpret exactly what is meant by this, it seems that this represents bodily cleanliness, perhaps related to occupation. So this might be a possible explanation for the diversity of risk within social class dependent on husband's occupation.

3.2.11 Others

A number of factors were identified that appeared protective against cervical cancer. These included dietary factors - especially high dietary intake of β-carotene (Vecchia et al., 1984) and vitamin A (Romney et al., 1981) - and previous history of diathermoelectro-coagulation of the cervix (DKG) (Vonka et al., 1984b, 56). On the other hand, use of the oral contraceptive pill (Stern et al., 1970) and cigarette smoking (Tokuhata, 1967; Winkelstein et al., 1984) both increased the risk of developing the disease.

In summary, a number of complex risk- and protective- factors were identified for cervical cancer. In general, these factors suggested that cervical cancer shared some aetiological similarities with

sexually transmitted diseases. However, the precise nature of the relationship between risk factors was extremely unclear, with no consensus about their status as either cause or covariate of cause. Multiple explanatory hypotheses were offered to accommodate these risk factors. This debate continued, even when mechanisms for cervical cancer had been suggested (Rotkin, 1967; Terris et al., 1967; Thomas, 1973; Kessler, 1976; Kessler, 1977).

3.3 Domain finding: HSV and cervical cancer

One explanation of the pattern of risk factors seen in cervical cancer was that the disease was caused by a transmissible agent, possibly a virus. While this hypothesis is mentioned in general terms in many of the publications identifying risk factors, the first case in which a specific agent is mentioned dates from 1966. This publication (Naib et al., 1966) suggested that cervical cancer might be caused by infection with HSV. In a study of about 40 000 cell samples from cervical screening tests in Atlanta, the authors demonstrated a strong correlation between HSV CPE and signs of malignant or pre-malignant changes of the cervical epithelium (Naib et al., 1966, 1026). They offered three plausible explanations for this phenomena. First, this apparent link might be nothing more than a coincidence. Second, HSV might have some predilection for malignant cells, and thus the correlation could be explainable in terms of an increased risk of contracting HSV in women with cervical cancer. Finally, HSV infections might be causally implicated in oncogenesis:

> . . . it can. . . be hypothesized that either a pre-existing cervical atypia invites a secondary viral infection or genital herpetic infection had some oncogenic potentiality
>
> (Naib et al., 1966, 1031)

This observation was, as the authors admitted, not completely novel. The characteristic CPE of herpesviruses had been previously detected in cervical screening samples. What was novel, however, was the possible causal role suggested for HSV. This idea quickly gained credence; subsequent authors noted that women with cervical cancer seemed to have a higher risk of being infected with HSV.

Causation of cervical cancer by HSV appeared to offer a highly plausible explanation of the known risk factors (Alexander, 1973; Kessler, 1976). First, the virus was known to be transmitted venereally. This offered a possible explanation of the apparent role of sex as an aetiologically significant factor in the development of the tumour. Second, the correlation between cervical dysplasia and HSV infection seemed to involve one particular type of the virus, HSV2, which is responsible for most cases of genital herpes infection (Cleator and Klapper, 2004a, 35). Thus an explanation could be formulated for the causal role in terms of host-cell tropism. Third, the virus could be isolated from smegma (Rawls et al., 1968a), and infections with HSV2 occurred at lower rates in circumcised males than uncircumcised ones (Parker and Banatvala, 1967, 212; 216). This could therefore offer a possible explanation of the apparently protective role offered by circumcision of sexual partners. This could in turn offer an explanation for the low rates of cervical cancer seen in members of faith groups who practiced ritual circumcision, especially if other ritual and cultural sexual practices also reduced the chance of contracting the virus (Alexander, 1973).

Explicable too was the role of youth at first sex, marriage or first pregnancy. If the virus were first transmitted to the women during adolescence, and the uterine cervix had a hypothetical greater sensitivity to any oncogenic effects of the virus at this age, it would be plausible that these risk factors could be explained by the actions of the virus (Rotkin, 1973).

If a multi-step model of oncogenesis, with HSV infection as a necessary component was hypothesised, then various other risk factors could be also explained. For example, the role of social class, diet and smoking could all be compatible with a causal mechanism driven by HSV infection.

This plausible role for HSV2 was reinforced by a number of apparently specific indicators of a causal mechanism. First, HSV2 and cervical cancer appeared to be serologically associated. Antibodies against HSV2 were present in prostitutes more than twice as often as in control populations (Rawls et al., 1969) and four times as often in women with cervical cancer than controls (Rawls et al., 1968b). Second, HSV was known to be capable of causing chromosomal abnormalities *in vitro* in both animal (Hampar and Ellison, 1963) and human cells (Stich et al., 1964). These abnormalities were similar to those known to be associated with many sorts of cancer. Third, it also became apparent that fragments of HSV DNA could be directly detected in cervical cancer cells (Frenkel et al., 1972), suggesting some specific role in pathogenesis. Finally, other herpes viruses were also implicated in other malignant diseases. In particular, when all this suggestive evidence was interpreted in the light of the recently discovered causal link between EBV and BL (see section 2.4ff), HSV became a highly plausible cause of cervical cancer. However, plausibility alone would not digest these risk-factors, and much effort was therefore focused on attempting to formulate a causal mechanistic link between HSV and cervical cancer, a selection of which I will detail in the next section.

3.4 Making mechanisms I: HSV (1966-1983)

By the early 1970s, HSV was generally accepted as the cause of cervical cancer. A survey of publications from the field at this time, for example those taken from the proceedings of the 1972 American Cancer Society conference '*Herpesvirus and cervical cancer*', demonstrate the existence of a thriving research program surrounding the HSV mechanism. Discussion of the nature of evidence for and against it, considerations of other possible causal agents, discussions of the similarities and differences between the case of cervical cancer and of other virally-caused tumours and discussions of the best way to continue investigations are all present. Two broad classes of hypothesis relating to herpes were in play. First, HSV and cervical cancer might both correlate with sex (Rawls et al., 1973, 1481). However, the finding that other sexually transmitted diseases did not correlate in such a way with cervical cancer led toward a second conclusion, which was that HSV and cervical cancer might be aetiologically associated (Rawls et al., 1973, 1482). This lack of correlation further undermined the position of other possible aetiologic agents considered. These had included *Trichomonas*, *Chlamydia*, and cytomegalovirus (CMV), while other agents that appeared to correlate with cervical cancer (for instance gonorrhoea, syphilis, and *Mycoplasma* spp.) had previously been ruled out as mere markers of sexual activity (Alexander, 1973). However: 'Whatever the evidence for the association of these agents with cervical cancer, none appears to be so likely a candidate for its cause as herpesvirus type 2.' (Alexander, 1973, 1485).

BL as a useful model system for cervical cancer was extensively considered, in combination with other animal models of disease, especially Marek's disease and *Herpes saimiri*.[5] These diseases offered researchers naturally occurring experimental models to understand sections of the HSV mechanism. However, unlike the very close parallels between the clinical manifestations of EBV infections and *Herpes saimiri* (Deinhardt, 1973), there was no such obvious role-model that closely modelled the whole disease process for cervical cancer and HSV. As one author commented:

> I do not think that they are models beyond what they show, namely, that some herpesviruses, quite different from HSV, can induce neoplasia under certain conditions. These conditions are very different, however, from the way in which a postulated oncogenic effect of HSV would take place.
>
> (Klein, 1973, 1558)

Given this lack of a suitable 'natural' disease model, researchers attempted to produce experimental disease models. However, most attempts at doing this appeared problematic. Early experimental attempts to induce tumours in animals by inoculation with various HSV types failed (Rapp and Falk, 1964). Later animal inoculation experiments too were equivocal, although some of them displayed a slight apparent effect of HSV2 inoculation on tumour development (e.g. Nahmias et al., 1970), although this work was complicated by the very high subject mortality caused by the lytic effects of HSV.

Similar experiments using *in vitro* cell cultures behaved very differently. For instance, Duff and Rapp, 1971 showed that cells *in vitro* were transformed by incubation with UV-inactivated HSV2. These transformed cells, known as the 333-8-9 line, continued to display HSV antigens (Duff and Rapp, 1971, 472), and were capable of inducing malignant tumour formation when inoculated into animal subjects (Duff and Rapp, 1971, 471). Similar findings occurred with HSV1 (Duff and Rapp, 1973). These findings were rapidly repeated by other groups using a variety of animal cell lines for both HSV1 and HSV2 (Kutinová et al., 1973; MacNab, 1974; Boyd and Orme, 1975; Duff and Rapp, 1975; Kimura et al., 1975). Fragments of HSV nucleic acids were also detectable in these cell lines (Collard et al., 1973; Frenkel et al., 1976) although it became apparent that this was an unstable property of the cells, as HSV DNA became rapidly undetectable when cell culture was prolonged (Minson et al., 1976).

The success of cell transformation *in vitro* led to a resurgence of attempts to directly induce cervical tumours in experimental animals by HSV exposure (Muñoz, 1973; Sever, 1973, 1510; Palmer et al., 1976; Fish et al., 1982). A further study also demonstrated the induction of cervical carcinoma, rather than atypia, using UV-inactivated HSV (Wentz et al., 1975). I will not talk about these (generally unsuccessful) attempts in detail, but I might speculate that the effects of publication bias makes it likely that there were even more unsuccessful attempts to induce cervical tumours using HSV than the literature reveals.

HSV2 nucleic acids were also recovered, in a single case, from cervical cancer tissue samples (Frenkel et al., 1972). However, this finding was not replicable, and it remained the case that HSV

[5] A herpesvirus affecting primates whose clinical manifestations include T-cell lymphomas. See section 1.1.

genetic material could generally only be detected from cells grown in *in vitro* cell cultures (Aurelian et al., 1971; Melnick et al., 1974). In order to strengthen this highly incomplete mechanism, the following suggestions were made for an ideal programme of ongoing research:

1. To improve HSV tests such that easy differentiation of HSV1 and 2 would be possible in practice[6]
2. To conduct prospective epidemiological studies, including recording data regarding subjects' hormonal status, details of their sexual partners and serial viral antibody titres.
3. To better understand the nature of the HSV life-cycle in the cell, including details of proteomics, genomics and mechanism of cellular transformation. This should include parallel work on similar conditions in an attempt to develop an animal model, including *Herpes saimiri*, *Herpes ateles*, Marek's disease, Lucké renal adenocarcinoma and EBV.
4. To better understand the disease immunology, including both a serological survey and an investigation of the possibility of immunotherapy.

(Deinhardt, 1973)

Much of this research on cells *in vitro* was inspired, in part, by the demonstration of EBV in cultured cells of BL (Duff and Rapp, 1971, 469; 476). One important suggestion for resolving these difficulties of causal inference was to carry out an intervention. If it could be shown that immunisation against HSV2 could reduce cervical cancer incidence, this would count as very strong evidence for the aetiological role of HSV in the pathogenesis of the disease:[7]

> Dr. Nahmias suggested that the proof concerning an etiological involvement of HSV-2 in cervical carcinoma may be obtained only by showing that immunization against HSV-2 has a preventive effect.

(Klein, 1973, 1561)

However, developing such an intervention never in fact came about.

3.5 Application I: HSV

3.5.1 Confirming the mechanism

Of primary importance in causation, given this absence of a specific causal mechanism, were statistical studies of HSV2 seroepidemiology in cervical cancer. Nearly all retrospective studies showed that patients with cervical cancer had higher anti-HSV antibody titres than control subjects, and that, for individuals with evidence of HSV infection, those with cervical cancer underwent infection earlier in

[6] Workable serological tests were eventually developed that were capable of discriminating HSV1 and 2 (Suchánková et al., 1984)

[7] Similarly strong evidence for causation arising from vaccination happened as part of the research programme linking hepatitis B virus (HBV) infection with the development of hepatocellular carcinoma (HCC). See section 4.5.2 for further details. Other authors have suggested interventions as proof of causation in cases such as these (Vonka, 2000; Bosch et al., 2002, 246). See also section 5.8

life (reviewed in Rawls et al., 1973). These antibodies appeared before the development of cervical dysplasia,[8] and their level reduced with its presumptive cure (Catalano and Johnson, 1971).

This work was supported by the specific correlations noted between cervical dysplasia and the presence of HSV CPE noted in individual histological examinations of cervical cancer samples, although not in cytological examination of cells from Papanicolaou smears (Naib et al., 1973). As no prospective studies were available at this point,[9] this work therefore relied on retrospective studies, a class of investigation subject to characteristic systematic errors.[10] However, this did not seem to be especially problematic in the case of cervical cancer research: matched breast cancer patients had similar recall of sexual behaviours (Melnick, conference discussion, cited Klein, 1973, 1561-2). Given this assumption of high quality, these retrospective studies indicated that detection of anti-HSV antibodies approximately correlated with cervical tumour burden (Thiry et al., 1974). This paralleled similar work in BL (Henle et al., 1973), but was complicated by experimental difficulties in serologically distinguishing HSV1 and HSV2 infections (Wildy, 1973; Plummer, 1973). This was especially pressing: infections with HSV1 were felt to be likely modifiers (by immune system modulation) of the clinical course of infections with HSV2 (Klein, 1973, 1561). For example, Adam et al. (1973) showed no significant difference in HSV antibodies between controls and cervical cancer or dysplasia groups until HSV1 and HSV2 were distinguished. Once this was done, the patient groups had higher levels of HSV2 antibodies than matched controls. Many similar studies were carried out (reviewed by Melnick and Adam, 1978, 52-3).

Other features for optimising retrospective case-control studies - in particular urging researchers towards matching on a small number of features - were discussed (Schneiderman and Levin, 1973). I will discuss this at more length in section 6.6.

While no trial of immunisation against HSV was conducted, other interventions seemed to provide support for the aetiological role of HSV. First amongst these was the finding that the use of barrier contraception reduced the incidence of cervical cancer. Although no interventional trial of this method appears to have been conducted, evidence from observational trials (e.g. Boyd and Doll, 1964, 430-1; Rotkin, 1973) suggested that cervical cancer patients had a lower rate of use of barrier contraception than matched control patients with other gynaecological disorders. This observation thus functioned as a natural experiment of barrier contraception.

3.5.2 Problems with HSV; alternate models

While both epidemiological evidence and suggestive mechanistic evidence was strong for HSV and cervical cancer, it did not appear possible to detect the HSV genome in cervical cancer cells. It seemed generally impossible to detect either HSV1 or HSV2 nucleic acids in cervical cancer biopsies (zur Hausen et al., 1974b), quite unlike the ease with which EBV nucleic acids could be detected in EBV-related tumours. This specific case may have been caused by the relatively low sensitivity of HSV tests at this

[8] Cervical dysplasia, more correctly known as cervical intraepithelial neoplasia, is the growth of abnormal cells on the cervix. It is a pre-malignant condition for cervical cancer.

[9] Later prospective seroepidemiological studies, such as Vonka et al., 1987, would reveal no such effect, as I discuss in section 3.5.2

[10] Retrospective studies are frequently criticised for their susceptibility to recall bias, in particular

time (Pagano, 1975, 219-22). But even when assay sensitivity was improved, HSV was still undetectable in either cervical dysplasia or cervical cancer (Cassai et al., 1981). When combined with the other problems facing the HSV mechanism, some authors began to seriously consider the role of papillomaviruses in cervical cancer (zur Hausen, 1975, 40-3).

Other models of oncogenesis were therefore suggested. First, could HSV be responsible for just one step of oncogenesis? zur Hausen (1982) suggested that cervical oncogenesis could be a multi-step process. Infection with one virus, in this case HSV, could initiate tumourigenesis. This tumour could then be maintained by infection with a second virus, in this case HPV. This 'hit and run' hypothesis could explain the serological role of HSV, while too explaining the failures of detection of the virus in tumour cells. This initiation of oncogenesis by HSV without leaving nucleic acid traces was experimentally plausible (Skinner, 1976; Schlehofer and zur Hausen, 1982), could be related to a partially confirmed genetic mechanism (Galloway and McDougall, 1983) and was shown to occur in cervical epithelial cells (DiPaolo et al., 1990).

More damning evidence would come from the failures of prospective studies. While a large prospective study of more that 10 000 women in Prague agreed with earlier retrospective risk factor epidemiology (Vonka et al., 1984b), no differences in HSV2 status appeared to exist between carefully matched control subjects and cancer cases (Vonka et al., 1984a; Krcmár et al., 1986).[11] Thus prospective studies did not seem to support a causal role for HSV in the development of the cancer. Similar results for HSV2 were generated in other studies, although one of them did detect a correlation between HPV CPE and the development of cervical dysplasia (Adam et al., 1985), while the second suggested a possible explanation for the higher levels of anti-HSV2 antibodies seen in women with cervical cancer (Lehtinen et al., 1992).[12]

Despite the suggestive evidence gathered, HSV remained a plausible, rather than a proven, cause of cervical tumours. There was a strong association between anti-HSV antibodies and the disease. These were increased in women with cervical cancer, cervical intraepithelial neoplasia and cervical dysplasia, when compared to matched controls. Comparisons of breast and cervical cancer patients suggested that this serological association was not just a function of recall bias. Finally, the expression of non-virion proteins in cervical cancer samples, and the apparent recovery of sections of HSV DNA from cervical cancer tissue, appeared to suggest a causal rather than a passenger role for HSV in the genesis of the disease. While this mechanism was highly incomplete, it was still generally accepted. I suspect that a major reason for this acceptability was its high degree of apparent plausibility, illustrated in its easy accommodation of the known risk factors.

[11] Each individual who developed cervical cancer was matched with up to four healthy controls on the basis of five factors. These were age, age at first sexual intercourse, number of sexual partners, smoking and history of cervical diathermy. I will return to this process of selecting factors in section 6.6. The Prague study was really the most important single piece of evidence against the theory that HSV caused cervical cancer. Interestingly, this apparently Popperian falsification of the theory was conducted at a time and place where the works of Popper were forbidden (Vonka, personal communication).

[12] They suggest that cervical cancer causes an increase in the quantities of HSV2 antigen presented to the immune system compared to the degree of presentation in women without cervical cancer. For a given severity of HSV2 infection, this leads to an apparent increase in antibody titres in cervical cancer patients.

3.6 Making mechanisms II: HPV (1983–)

3.6.1 Historical background

Despite the plausible aetiological role of HSV2, various other mechanisms also appeared to offer alternate accounts of cervical cancer causation. Foremost among these was HPV. It initially seemed a much less plausible cause of the disease than HSV. While herpes viruses were implicated in a wide variety of malignant disease states, less such supportive evidence existed for HPV. There was some, however, which I will briefly review below (also reviewed by zur Hausen, 1996a; zur Hausen, 1996b). In brief, the factors supporting a causal role for papillomaviruses in the development of cervical cancer were these. Papillomaviruses were known to cause tumours in humans. As their name suggests, by far the most common type of papillomavirus-induced tumour was the benign skin papilloma (wart). Some malignant conditions were also thought to be related to HPV infection, for instance EV, as I will discuss in section 3.6.1.2. Papillomaviruses were also responsible for a range of benign and malignant diseases of non-human animals. I will discuss two examples in section 3.6.1.3, BPV and CRPV.

3.6.1.1 Warts

Warts (benign skin papilloma) are common, benign skin tumours. They exist in a variety of forms, including the common skin wart (simple papilloma), verruca (plantar warts), genital and laryngeal. They have long been suspected to have a transmissible aetiology. For example, warts have been transmitted between a variety of anatomical sites by a plethora of means, both experimentally and accidentally (reviewed by Rowson and Mahy, 1967). Cases were reported where warts were transmitted through the inoculation of ultra-filtered material, suggesting that their aetiologic agent was sub-cellular in size (Wile and Kingery, 1919; Kingery, 1921). The apparently common nature of transmissibility, a suspected correlation between the occurrence of different types in individual cases, and histological similarities between types of wart (Findlay, 1930, 254) suggested a common cause for common, plantar, laryngeal (Ullman, 1923) and genital (Wilson, 1937) warts. This unitary theory of wart causation (Oriel, 1971, 2) began in the 1920s and was not seriously challenged until the 1970s.

As warts were transmissible by CFF, a virus seemed a highly likely cause. Research into this suggestion followed a similar course to the other examples of viral oncogenesis discussed above. For instance, the use of preventative intervention was suggested as a useful way of determining causation (Biberstein, 1944). However, the viral hypothesis really took off with the wide-spread availability of the electron microscope,[13] which revealed considerable evidence for the presence of viruses in wart cells. For instance, crystalline aggregates of VLPs were demonstrable (Strauss and Shaw, 1949). These particles had a characteristic morphology (Melnick et al., 1952), which appeared consistent between samples (Bunting, 1953; Reagan et al., 1954), although there was some minor disagreement about their diameter (Siegel, 1960). Inclusions bodies, suggesting active viral replication, were also found in wart material (Lyell and Miles, 1951), some types of which seemed to be correlated with the microscopic presence of aggregates of viral particles when examined microscopically (Almeida et al., 1962). Similar co-location results were produced by the use of immunofluorescence techniques, using antigens derived

[13] see section 1.1

from these virus particles (Walter et al., 1965).

All this implicated these virus particles in the causation of warts. However, it proved very difficult to grow these viruses in culture systems. While Bivins, 1953 reported presumptive success at growing an agent derived from common warts in fertilized hens eggs, this finding was not replicable by other authors (Felsher, 1947; Fischer, 1953; Siegel and Novy, 1955). This led to suggestions that the successfully cultivated virus in the Bivin's case was a contaminant, possibly canary pox virus (Siegel, 1956).

Only in human cell cultures did it seem possible to cultivate the virus.[14] Successes were reported using material derived from both genital (Morgan and Balduzzi, 1964) and common warts, with the production of large quantities of viral particles (Oroszlan and Rich, 1964). Cultures of wart viruses in human cell cultures appeared to lead to morphological and behavioural changes in the culture cells (Noyes, 1965). HPV was therefore thought to be non-culturable *in vitro* (Butel, 1972).

This highly specific viral tropism represented a substantial difficulty for researchers working on the wart virus. As research continued, it became apparent that unitary theory of wart causation was flawed. Despite the apparent similarities between wart viruses in terms of clinical features, transmissibility, viral morphology and culturability, warts at different anatomical sites had subtle, but important, differences. Taking the case of genital warts, the natural history seemed distinct from common cutaneous warts (Oriel, 1971). Other findings such as the correlation between multiple types of warts in particular patients, also appeared spurious (Oriel, 1971). While papillomavirus-like particles were sometimes visible in genital warts under electron microscopy, in common with other types of warts, (Dunn and Ogilvie, 1968; Oriel and Almeida, 1970), they generally appeared at a lower frequency. In conclusion, and importantly for our story, it also appeared that wart viruses with different host tropisms were immunologically distinct (Almeida et al., 1969).

3.6.1.2 Epidermodysplasia Verruciformis

EV is a rare, familial skin disease, which was first described by Lewandowsky and Lutz, 1922. It is characterised by lifelong eruptions of large, transmissible skin lesions resembling planar warts. These lesions, unlike true warts, have a high risk of malignant transformation, with about 25% of cases eventually developing some skin malignancy (Jablonska et al., 1972). Papillomavirus particles are detectable within the lesions in a similar way to other warts (Delescluse et al., 1972), although they become undetectable if malignant transformation occurs (Ruiter and van Mullem, 1970).

This condition was important for papillomavirus research in two ways. First, it implicated papillomaviruses in a malignant disease of humans. Second, once type variability of the papillomaviruses was discovered in warts, EV became an important second source for discovering new papillomavirus types with differing pathological proclivities (for instance, Orth et al., 1978b).

[14] One success was apparently obtained in animal cell culture: Mendelson and Kligman, 1961 reported the isolation of the wart virus using cultures of MK (monkey kidney) cells. This virus, when inoculated into human volunteers, led to the development of warts at the inoculation sites. Interestingly, the authors note that this research fulfilled Koch's postulates for the agent. This seems unlikely owing to the requirement for growth in pure culture.

3.6.1.3 Animal tumours and papillomaviruses

Two types of papillomavirus found in non-human animals were particularly important to this story.

BPV, a member of papillomavirus superfamily C, was known to cause cutaneous papillomas in cattle in a similar fashion to HPV in humans. However, this agent caused malignant bladder tumours when injected into cow bladder mucosa (Olson et al., 1959), and was capable of causing similar cellular transformation *in vitro* (Black et al., 1963). This suggested that individual papillomaviruses types might have a 'natural' pathological effect at one anatomical site (so they might cause warts if they infected the skin) but might also have an alternate effect if introduced into a different site (so they might cause cancer if they were to infect the cervix). Like HPV, while it was cultivatable in tissue culture using cells of its tropic host, it was extremely difficult to induce the virus to affect cells of any other species (Boiron et al., 1964).

The second, CRPV,[15] a member of papillomavirus supergroup E, was the first animal papillomavirus identified. Its clinical manifestations are large, cutaneous fibrous tumours, similar to large warts, that occur in wild cottontail rabbits (*Sylvilagus floridanus*). It was shown to be transmissible by cellular inoculation (Shope, 1932) and by inoculation with cell-free filtrates (Shope and Hurst, 1933) to both other cottontail rabbits and to common domestic rabbits (*Oryctolagus* spp.). Malignant change in these papillomas was later observed (Rous and Beard, 1934; Syverton and Berry, 1935), although this may have been a feature of the disease only when induced in *Oryctolagus* spp. (Rous and Beard, 1935). The tumours also appeared to induce serum immunity to the agent (Rous and Beard, 1935, 541-2). Malignant transformation appeared to be synergistically potentiated by the addition of chemical carcinogens (Rous and Kidd, 1938; Rous and Friedewald, 1944). Viruses were detected in these papilloma (Moore et al., 1959), and demonstrated to be papillomavirus by electron microscopy (Stone et al., 1959).

CRPV has been extremely important in other ways to recent papillomavirus research as a model-system of papillomavirus action. While the details are beyond this piece, an excellent technical review is Breitburd et al., 1997.

3.6.2 HPV types

So working with papillomaviruses presented significant challenges to researchers. By the early 1970s, papillomaviruses had become a troubling puzzle. Much evidence supported the idea that various clinical diseases - warts, EV, bovine papillomas and cottontail rabbit fibromas - were apparently related to infection with papillomaviruses. Electron microscopy of material derived from different types of warts (human skin, human genital and various kinds of warts in animals) revealed strong morphological similarities between the various virus particles (Oriel and Almeida, 1970). This morphological similarity seemed to fit in with the similarities of the clinical course of the disease seen during experimental or accidental wart transmission experiments (as discussed in section 3.6.1.1). Together, these two features suggested that papillomaviruses could cause both benign and malignant diseases, and were transmissible.

Yet this support was more equivocal than first appearances would suggest. For one, the different known papillomaviruses, despite having similar gross morphologies and genome lengths (Crawford and

[15] also known as the Shope papillomavirus

Crawford, 1963), were antigenically distinct (Bouvier et al., 1966) and their genomes showed a good deal of variance when analysed in terms of base composition (Crawford and Crawford, 1963, 262). Both immuno-electron microscopy and RNA-hybridisation experiments revealed a significant degree of difference between the papillomaviruses associated with different diseases, with material derived from one type of disease appearing not to share immunological nor genetic characteristics with material derived from other types (Almeida et al., 1969; zur Hausen et al., 1974a; Delap et al., 1976).[16] There were important epidemiological questions too. Different papillomavirus diseases in humans (warts, genital warts or laryngeal warts) characteristically occurred at different ages and in different populations. This finding too seemed to count against a strongly unified cause for these diseases. For instance, the results of transplantation experiments on skin and genital warts (where warts were removed from one individual and implanted on another, or where a wart was moved between anatomical sites on one subject) were, at best, highly equivocal (see Oriel and Almeida, 1970, 37 for a review of these findings), suggesting that even if papillomaviruses were implicated in the causation of these diseases, other unknown causal factors were also important. Even more puzzling was the failure of the development of molecular methods to detect these viruses. For instance, it did not appear possible to detect HPV DNA in warts, despite all the other supporting evidence for their causal role (zur Hausen et al., 1974a).

Were HPV a group of related viruses? Could this explain the different clinical and experimental properties of the (visually identical) viruses? Part of the puzzle was resolved when researchers examined in detail the genomes of papillomaviruses from different types of wart material. It rapidly became apparent that a number of diverse HPV types were present. This difference applied too even between different samples from warts of similar type-for instance Gissmann and zur Hausen, 1976, who reported two apparently different papillomavirus types by restriction endonuclease digestion from plantar wart material. This was not an isolated finding. Favre et al., 1975, during work to map the genome of HPV, discovered a degree of genetic variation between types.

What was the significance of this finding for cervical cancer research? It led to the first serious suggestion that, despite the lack of detection of the virus in cervical cancer cells, HPV could still be the cause of cervical cancer (zur Hausen, 1976). zur Hausen commented:

> The condyloma agent has been entirely neglected thus far in all epidemiological and serological studies relating . . . to cervical . . . carcinomas. This is particularly unusual in view of the localization of genital warts, their mode of venereal transmission, the number of reports on malignant transition, and the presence of an agent belonging to a well-characterized group of oncogenic DNA viruses.
>
> (zur Hausen, 1976, 794)

In fact, zur Hausen was later to suggest that three features of HPV were particularly important in the decision to conduct a detailed investigation of their role in cervical cancer (zur Hausen, 1983, 317-8). These were, first, the failure of HSV detection in cervical cancer (as discussed in section 3.5.2), the

[16] For a contemporary review of the details of techniques used to detect HPV in a variety of settings see Koutsky et al., 1988.

venereal transmission of HPV (as suggested by the sexual transmission of genital warts) and the ability, in rare cases, of genital warts to undergo malignant transformation (Friedberg and Serlin, 1963).[17]

But this would all depend upon demonstrating distinct types of HPV. Gissmann et al. (1977) described the detection of four types of HPV (HPV 1, 2, 3 and 4)[18] in skin wart material. While HPV 1-3 appeared closely related to each other when examined by a variety of techniques (by restriction endonuclease digestion and electrophoresis, by RNA hybridisation and by immunological methods), HPV 4 seemed to be an altogether distinct virus. These differences seemed to correlate with histopathological differences too. Wart samples with low numbers of virus particles tending to be infected with HPV 4 when examined microscopically. Yet the virus particles themselves were indistinguishable from each other (Gissmann et al., 1977, 577). More interestingly, of the 50 samples of skin warts examined, 12 did not show evidence of infection with any of the four types. Neither were they detected in genital or laryngeal wart samples (Gissmann et al., 1977, 579). Were there more types to be found?

A further two types of HPV were provisionally described by Orth et al., 1977. One, HPV 1, which shared features with HPV 1-3 identified by Gissmann et al., 1977 was detected in a plantar wart. The second was a novel type from a common skin wart, which became known as HPV 2. Again, they had identical microscopic morphologies and similar-size genomes, produced different fragments on restriction-endonuclease digestion and lacked homology on RNA-DNA hybridisation. While HPV 2 did not behave similarly to any of Gissman's 4 types, Orth's HPV 1 was similar to Gissman's HPV 1-3. Hybridisation experiments with the new HPV 2 material against pooled HPV isolates revealed hybridisation, leading to the conclusion that HPV 2 was a cause of warts in general.

At much the same time, isolates from EV were seen to be immunologically distinct from wart types found in plantar and skin warts (Pass et al., 1977). Two new types were described (Orth et al., 1978b), designated JD HPV and JK HPV. Again, these types had similar electron microscopy and genome characteristics, but were distinct from each other and from HPV 1 and HPV 2 by hybridisation and immunology. Again, the different types were associated with clinical differences: JD was discovered in a flat, wart-like EV lesion, while JK was found in red, plaque-like lesions, which had a tendency towards malignant transformation (Orth et al., 1978b, 1541). This suspicion was reinforced by a later clinical study showing just that (Orth et al., 1979).

As might be expected, HPV types from genital warts had similar morphological and genomic properties to those from other sources, but showed little sequence homology with any other HPV types (Orth et al., 1978a). However, genital warts presented an especially challenging research target, owing to the relative paucity of viral particles present in them. It was not until Gissmann and zur Hausen, 1980 that extraction of a genital isolate was achieved, made possible by a sample containing exceptionally high levels of virus. The main development here was the extraction of DNA from a novel type, designated

[17] zur Hausen was, for many years, almost the sole supporter of the theory that papillomaviruses played a causal role in cervical cancer. This prolonged involvement in a research programme that was unsuccessful for so long has been lately recognised by the Nobel committee (zur Hausen, 2008).

[18] The naming of HPV types became extremely confusing. In the following passage I will use the terminology used in the original publications. However, the known types were reclassified in July 1978 (Coggin and zur Hausen, 1979) into HPV 1-5 by a combination of sequence homology and serology. See table 3.1.

HPV-6, isolated from genital warts. Again, it seemed to share no homology with HPV 1-5. However, all samples of genital wart material tested against the virus derived from a single sample seemed to show homology. It was thus suggested that genital warts, contra the other types of papillomas, might display a high degree of causal homogeneity. The paper also ends with a return to considering malignant disease:

> Characterization of these agents should provide the basis for a further analysis of their possible involvement in human malignant disease.
>
> (Gissmann and zur Hausen, 1980, 608)

Type isolation proceeded rapidly, with HPV-7 isolated from cutaneous warts from meat workers (Orth et al., 1981; Ostrow et al., 1981). This virus was distinct from BPV. Further isolates were made from patients with EV (Pfister et al., 1981), and with laryngeal papilloma (Gissmann et al., 1982a).

HPV-6 material was cloned, and used as a DNA probe (de Villiers et al., 1981; de Villiers et al., 1981) in order to look for papillomavirus material in other samples (zur Hausen, 1983) In summary, 6 malignant genital warts (Buschke-Löwstein tumours) contained HPV-6 or -11 DNA (Gissmann et al., 1982b). Importantly for our story, five of 27 cervical cancer biopsies examined contained HPV-11 (Gissmann et al., 1983). Similar findings were made by other groups (Green et al., 1982; Okagaki et al., 1983). While this evidence was suggestive of a role for HPV in cervical cancer, it was not until the discovery and cloning of HPV-16 (Dürst et al., 1983) that HPV could be detected in most cervical cancer tissue samples. The story of the discovery is interesting. First, a number of cervical cancer tissue samples were screened using HPV-11 DNA under nonstringent conditions. One of these (WV 2916) contained a papillomavirus sequence which hybridised with HPV-11. However, under conditions of high stringency this sample failed to hybridise with known HPV types. When it was analysed by restriction endonuclease digestion, it revealed an unusual cleavage pattern, suggesting that this was a novel type. Thus a new papillomavirus was detected in cervical cancer tissue. When this new sequence was used to probe 18 German cervical cancer samples, it hybridized with high stringency with 11 of them. In comparison, when screened against genital wart material, it was present in only two of 33 samples used. There also seemed to be a degree of geographical variation, with the sequence being detectable in only eight of 23 Kenyan and Brazilian cervical cancer samples screened. Thus this isolation seemed to be a very likely cause of cervical cancer, in European women at least.

One of these negative Kenyan samples was shortly shown to contain a second novel type, HPV-18, by similar means to that in HPV-16 (Boshart et al., 1984). As might be expected, fewer German (2/13) or Brazilian (9/36) samples were positive. Related DNA was also found in a number of existing human cervical cancer cell lines, including HeLa, KB and C4-1.[19] Taken together, HPV-16 and -18 were present in most tested cervical cancer samples. But what was the pathogenic mechanism?

[19]The detection of HPV-18 in HeLa cells is, perhaps, unsurprising. These cells, named for the patient from whom they were harvested (**He**nrietta **La**cks), were derived from a cervical tumour in 1951 (Skloot, 2010). They are the most commonly encountered human cell in research practice. The induction of this tumour by HPV-18, and its continuing presence in these cells, has since been confirmed (Macville et al., 1999).

Publication date	Publication	July 1978 reclassification	Original type designation	Source material
Dec 1975	Favre et al., 1975	HPV-3	?	?
Feb 1977	Gissmann et al., 1977	HPV-1a	HPV 1	Skin warts
Feb 1977	Gissmann et al., 1977	HPV-1b	HPV 2	Skin warts
Feb 1977	Gissmann et al., 1977	HPV-1c	HPV 3	Skin warts
Feb 1977	Gissmann et al., 1977	HPV-4	HPV 4	Skin warts
Oct 1977	Orth et al., 1977	HPV-1	HPV 1	Plantar wart
Oct 1977	Orth et al., 1977	HPV-2	HPV 2 (ML)	Skin warts
Mar 1978	Orth et al., 1978b	HPV-3	HPV 3 (J.D)	EV, flat lesions
Mar 1978	Orth et al., 1978b	HPV-5	HPV 4 (JK)	EV, red-spot lesions

Table 3.1: Summary of early types of HPV, showing original designations and post-classification designations as per Coggin and zur Hausen, 1979.

3.6.3 HPV-mediated oncogenesis

While I won't attempt a detailed review of the mechanism by which HPV causes cervical cancer, it is worth discussing a few of the major developments here.[20] First, it was found that cloned HPV DNA alone was capable of transforming cells *in vitro*. While this was first demonstrated with HPV types not thought to be implicated in cervical cancer (Watts et al., 1984), it was later shown to be possible with HPV-16 (Yasumoto et al., 1986). This ability of HPV-16 to immortalize animal cells *in vitro* was found to extend to human cells of various types (Dürst et al., 1987; Pirisi et al., 1987), including cells of the cervical epithelium (Chen et al., 1993). This process of transformation involved incorporation of the HPV genome into the host genome and its amplification (Schwarz et al., 1985), possibly explaining the high levels of HPV DNA detectable in human cervical cancer cell lines. HPV integration was found to be the case in most human cervical cancer cell lines (Yee et al., 1985).

RNA analysis revealed the expression of a number of non-structural viral proteins in these immortalized cells.[21] These 'early' viral proteins appeared causally important in cellular transformation (Kaur and McDougall, 1988). In particular, the E6 and E7 proteins appeared to be responsible for modifying cell cycle regulation (von Knebel Doeberitz et al., 1988) through a number of specific mechanisms. Acting together, HPV-18 E6 and E7 proteins were found to be necessary and sufficient for cell transformation, *in vitro*, at least (Münger et al., 1989).

E6 is necessarily responsible for the maintenance of tumours (von Knebel Doeberitz et al., 1994). However, acting alone, it also appeared capable of inducing immortalization in some cell-types (Band et al., 1991). It causes cell-cycle deregulation by p53 degradation (Scheffner et al., 1990). E7, though, acts through interactions with pRb (Dyson et al., 1989), leading to cell-cycle deregulation via disruption of the actions of transcription factor E2F (Chellappan et al., 1992). Thus, E7 acts as the initiator of cell immortalization (Halbert et al., 1991).

This process of HPV integration, early protein expression and disruption of cell-cycle regulation was suggested as also being capable of inducing malignant transformation, rather than immortalization. For instance, *in vitro* studies of keratinocytes immortalized by HPV showed that they might occasionally became malignant upon prolonged passage (Hurlin et al., 1991; Pecoraro et al., 1991). However, it is generally the case that HPV is seen as an agent capable of producing cell immortalization, rather than malignant transformation *in vitro*, and most researchers use another transforming agent (such as activated H-ras) to produce maligant cells experimentally (Vonka, personal communication).

[20] I can recommend the following reviews of the subject: zur Hausen, 1986; zur Hausen, 1991; zur Hausen, 1996a; zur Hausen, 1996b; zur Hausen, 2000; McCance, 2004.

[21] See also Moody and Laimins, 2010 for a review of the current state of knowledge on the manner in which these early proteins act to transform cells.

3.7 Application II: HPV

3.7.1 Testing the mechanism

By this point, almost all the evidence implicating HPV in the development of cervical cancer had arisen from laboratory studies (zur Hausen, 1989a, 1680; zur Hausen, 1989b).[22] However, in the early 1990s, work was conducted that attempted to reconcile the causal mechanism identified in the laboratory with the known risk factors for cervical cancer. The first step was to show general applicability. This occurred in two ways. First, von Knebel Doeberitz et al., 1992 noted that the promotion of the expression of E6 and E7 *in vivo* led to increased tumour growth, while their inhibition led to regression. So HPV appeared to act in a similar way in either experimental systems or in whole organisms.

Second, population studies were conducted to study the relationship between HPV, previously identified cervical cancer risk factors and cervical cancer itself. Bosch et al., 1992, a case-control study between Spain and Colombia suggested, first, that HPV infection (as assessed by polymerase chain reaction (PCR)[23] in exfoliated cervical epithelial cells) was a strong risk factor for cervical cancer. Further analysis of the same population sample suggested that the link between HPV-16, -18, -31, -33 and -35 was likely to be causal, on the grounds primarily of the strength of the association.[24] Second, this study suggested that some risk factors, in particular promiscuity, could be explained largely in terms of risk of HPV exposure. However, some risk factors, while plausibly accounted for in terms of modifying HPV exposure risk, could not be accounted for in terms of HPV in this study.

Similar findings occurred in other epidemiological surveys, showing a strong association between HPV and cervical cancer across populations. More than 90% of cervical cancers were HPV positive (Bosch et al., 1995), and adjustment for other risk factors suggested that in a very high proportion of cases it was infection with HPV that was causal (Schiffman et al., 1993). As previously suspected,[25] while the precise prevalence of HPV types varied geographically (Bosch et al., 1995), in most cases HPV-16 and -18 appeared to be the most important in terms of cervical cancer causation.

---

[22] 'Interestingly and in remarkable contrast to other viruses linked to human cancer... most of the evidence originated from laboratory studies revealing transforming activity of these viruses, a specific mode of integration and of genetic activity in cancer cells, and the role of viral genes in the maintenance of the proliferative phenotype in cells of HPV-positive cervical carcinoma lines. To a much lesser degree, epidemiologic approaches contributed to our present state of knowledge, whereas seroepidemiology virtually contributed not at all' (zur Hausen, 1989a, 1680).

[23] PCR entered general research contexts in the early 1990s, offering in this case a relatively inexpensive and highly sensitive means of detecting HPV types far superior to other techniques (Bauer et al., 1991).

[24] 'When the results... were combined to better define clearly negative and clearly positive specimens an extremely strong association emerged, OR = 146.6 (35.5-606.4). Odds ratios of this strength are extremely rare in cancer epidemiology. The strength of this association points towards a causal association and there is at present no experimental evidence suggesting that this association could be explained either by enhanced susceptibility of cancer cells to HPV infection or by enhanced HPV-DNA detectability in cancer cells.' (Muñoz et al., 1992, 747). In fact, as a recent study suggests, this may well be a substantial underestimate of the OR associated with high-risk HPV types. For types -16 and -18 respectively, the odds ratios are 2770 (95% CI 1050-7320) and 950 (95% CI 330-2740), while the OR for the other oncogenic types are in general much lower (between about 20 and 400) in the same population (Powell et al., 2010).

[25] e.g. Dürst et al., 1983; Boshart et al., 1984

3.7.2 Sharp observations, sharp interventions

While I do not want to discuss the epidemiology of cervical cancer at any more length, I do think that it is worth discussing a couple of recent research themes that arose from it, which were significant in their own right as causal evidence. The first was using HPV status as a means of assessing cervical cancer risk, the second is the use of vaccination against HPV as a preventative and (potentially) therapeutic tool against cervical cancer (Vonka and Hamsíková, 2007).

Testing for HPV either as a replacement or as an adjunct for conventional cervical cytology has been suggested (Mayrand et al., 2006) and attempted in trial contexts, where it has been found to be more sensitive that cervical cytology alone (Mayrand et al., 2007). This appears to be the case using a variety of HPV testing protocols (Gravitt et al., 2008). However, this great sensitivity may well be problematic. For example, while pre-cancerous lesions likely to progress to malignancy will be detected by HPV testing, so will precancerous lesions likely to spontaneously regress (Ronco et al., 2008). The heterogeneity between HPV types is also a problem for sequence-based tests. Each system is designed to test for a certain range of HPV types. So depending upon the exact test used, different types may or may not be detected (Klug et al., 2008), complicating epidemiology and practice.

Successful vaccination programmes have been suggested as useful evidence for causation, as I discuss in sections 4.5.2 and 5.8. The majority of effort in this case is concerned with preventative, rather than therapeutic vaccination.[26] Two HPV vaccines have recently entered clinical use. Gardasil® (Merck & Co.), which is a quadrivalent vaccine active against HPV-16, -18, -6 (genital warts) and -11 (laryngeal warts) (Villa, 2007) and Cervarix® (GSK), a bivalent vaccine against HPV-16 and -18. Both are based upon recombinant virus-like particles of the various L1 major capsid proteins. While safe and apparently effective in terms of preventing HPV infection, they are relatively expensive, with a basic NHS cost of approximately £250 per course (Joint Formulary Committee, 2009). Vaccination programmes against HPV have therefore begun in much of the developed world (Senior, 2008), but costs have restricted the implemention of these programmes in the developing world (Anonymous, 2008).

Current vaccination research focuses on producing vaccines with activity against greater numbers of HPV types. In particular, attempts are being made to produce recombinant, chimeric, VLPs that include fragments of the L2 capsid protein, which provides a higher degree of immune system cross-reactivity (Kondo et al., 2008).

[26] Although there are currently attempts being made to use vaccines in a therapeutic way - for instance, Hung et al., 2008 and Massa et al., 2008.

Chapter 4

The Russo-Williamson Thesis

4.1 Introduction

In their 2007 paper, Russo and Williamson promote a thesis on the nature of causality in medicine. Causes, they suggest, are monistic. That is, there is only one sort of causality in medicine. This causal monism, though, depends upon evidential pluralism. More specifically, they claim that both mechanistic and probabilistic evidence are typically required to support causal claims in medicine. Thus, they claim that causes in medicine are metaphysically monistic, but evidentially pluralistic. This is the essence of their claim, and it will form the basis of this chapter. I intend to show that, with certain modifications, the Russo-Williamson thesis (RWT) can become operationally useful in examining causal claims from the history of medicine. This programme of creating an historically motivated version of the RWT will form the basis for the subsequent parts of this thesis.[1] Thus, the job of this chapter is to outline Russo and Williamson's arguments accurately, to examine them critically, and then to suggest a number of objections and counterexamples. By means of introducing the remainder of this thesis, I will conclude this chapter with a sketch of how these objections might be overcome. This chapter is, therefore, largely a review of the RWT and a series of problem statements. It will assume the following structure. First, in section 4.2, I will outline the claims that Russo and Williamson make. At the outset, I suggest that most of their claims are correct. But even here, I will suggest that certain aspects of their account are likely to be problematic when we try to understand and evaluate real, medical, complex causes. I will also outline a number of philosophical objections to the RWT in section 4.3.

Next, what light can the examination of detailed historical cases shed on the RWT? In section 4.4, I outline the reasons why I regard their thesis as a practical, historical proposition. This section will be based upon the development of a causal claim linking EBV infection with BL from chapter 2.

I conclude this chapter by introducing three *prima facie* historical counterexamples to the RWT. The first of these deals with the history of discovery of the causation of McArdle disease. This appears to be a case of robust causal inference made in the absence of probabilistic evidence. The second of these, the discovery of the role played by chronic HBV infection in the aetiology of primary liver cancer, appears to be a case where mechanistic evidence was lacking, and probabilistic evidence alone seemed sufficient

[1] Russo and Williamson employ some brief historical examples too: on smoking and lung cancer, the Semmelweis case, cholera, *Helicobacter pylori* and stomach ulcers. The historical side of their thesis will be developed further in what follows.

to support a sound causal claim. The third case, based upon the discussion of HSV and cervical cancer in sections 3.3-3.5, relates the development of a faulty causal claim which was supported by apparently high-quality probabilistic and mechanistic evidence.

How can we save the RWT in the light of these difficulties? The cases and counterexamples in this chapter will form the basis for the necessary developments. First, I will outline my position on mechanisms in chapter 5. I will then use features of the historical cases to develop an account of evidence for causality in chapter 6. This account is intended to show that Russo and Williamson's mechanistic/probabilistic distinction is historically and conceptually tenable. Here, I will suggest that we can make Russo and Williamson's differentiation between types of evidence for causality. Finally, in chapter 7, I will give an account of the nature of integration of evidence in the examples. Briefly, I suggest that mechanisms provide a means by which we can address the reference-class problem when examining our probabilistic evidence. We take the (likely) significant causal landmarks identified during the process of mechanism construction, and partition our probabilistic evidence using them. Thus, probabilities tell us about the pertinence of a causal mechanism to a given population, while mechanisms tell us about the manner in which an effect can come about in that same population.

With these developments in place, the RWT can accommodate both our examples and our counterexamples. But I also suggest that my account of the manner in which evidence is integrated is not only a feature of successful causal claims, but that it also represents an important epistemic tool that can provide us with the means to assess new or controversial claims of causation in medicine. But that is for much later in the thesis. For now, I will simply introduce the RWT, outline a range of philosophical objections to it, briefly illustrate some of its features by reference to the EBV example and set out the three historical counterexamples.

4.2 The Russo-Williamson Thesis

Russo and Williamson begin by making the claim that medicine requires the notion of causality. Medical researchers, they argue, look for causes. This opposes recent claims that such causal talk is 'redundant and misleading' (Lipton and Odegaard, 2005, 6), and is perfectly replaceable by non-causal terminology. Thus the employment of weakened causal terminology, such as 'risk factor' or 'determinant', in the medical literature does not reflect an underlying reduction in the importance of causation in medical practice (Russo and Williamson, 2007, 157-8). I think this claim is generally correct. Certainly, the EBV and HPV cases seem to support an historical case for the centrality of causality to medical research. So Russo and Williamson's first claim is that:

> It is quite uncontroversial that the health sciences look for causes, namely for causes of disease and for effective treatments.
>
> (Russo and Williamson, 2007, 157)

We might, for the sake of clarity, précis the first part of this quotation as:

RW1. Medicine requires the notion of cause

This quotation also contains a second claim, and one which is supported by the roles played by causes in the two historical cases. This is that medicine seeks causes that fall into two broad classes. These are aetiological causes and therapeutic causes (Russo and Williamson, 2007, 157). Aetiological causes are those that are concerned with the cause of diseases, while therapeutic causes deal with the manner in which a disease may be prevented or cured. This simple distinction is of profound importance. In general, it is simply not the case that therapeutic interventions always consist of preventing aetiological causes. For example, therapy for BL does not involve preventing infection with EBV. Instead, it involves the use of chemotherapeutic drugs which act by preferentially killing tumour cells. While a fuller account of the distinction and relationship between aetiological and therapeutic causes will have to wait for section 5.8, this claim seems historically well supported. Although the bulk of the EBV and HPV cases deal with the process of discovering aetiological causes, there are distinct hints of subsequent work directed at finding therapeutic causes apparent in both cases. So Russo and Williamson's second, historically supported claim is:

RW2. Two sorts of causal relationship - aetiological or therapeutic - are important in medicine.

However, Russo and Williamson then go on to formulate a slightly different type of distinction, one which differentiates the *reasons* that researchers hunt causes, rather than (as with **RW2**) distinguishing between types of cause. Medicine, they argue, seeks causes for both cognitive and action-oriented goals:

> On the one hand, the health sciences pursue a cognitive goal, detecting causal factors and identifying mechanisms of disease... on the other hand, they pursue an action-oriented goal, informing policies and guiding early diagnosis or treatment on the basis of causal knowledge...
>
> (Russo and Williamson, 2007, 157)

The cognitive goals for which causes are sought are largely encapsulated by considering the role of causes in forming good explanations. Action-oriented goals, on the other hand, are largely used to ground clinical medical practice - for example, in guiding policy, diagnosis and intervention. Now this distinction divides along a different axis from the aetiological/therapeutic distinction of **RW2**. It is not the case that there is a simple correspondence between cognitive reasons for seeking causes and finding aetiological causes or between seeking for action-oriented reasons and finding therapeutic causes. It is easy to imagine a cognitive goal for which a therapeutic type of cause is sought (for instance, by seeking to explain the success of a therapeutic agent). Likewise, we might imagine many action-oriented reasons for attempting to discover the aetiological relationships at work in a particular disease. Russo and Williamson ground this intentional distinction in a differentiation between the sorts of epistemic practices to which causal relationships can be put. Specifically, they suggest that the first of these functions reveals a potentially explanatory aspect to causal relationships, while the second goal reveals an inferential component.

RW3. Causes are sought for either cognitive or action-oriented reasons

As with **RW2**, this distinction appears logically very important. However, unlike **RW2**, I have doubts about the historical support that can be provided for it. Particularly, distinctions of this sort appear lacking in the viral oncogenesis cases. If we interpret **RW3** in a strong sense, where the different reasons for seeking causes reflects an issue of distinct agency, the following dilemma results from this lack of historical support. Either Russo and Williamson are wrong about this distinction; or we must suggest that the historical literature does not accurately reflect the way that causes are sought.

The first horn appears logically problematic. Good causes do appear to provide a rational licence for both these distinct sorts of practice. It therefore appears incorrect to suggest that this distinction is utterly baseless. On the other hand, accepting the second horn appears equally difficult for any analysis of causality that is grounded in (historical) medical practice. So some sort of alternative strategy seems appropriate here. My suggestion is this: that the cognitive/action-oriented distinction really just represents the uses to which causal relationships are put, rather than being any sort of issue of distinct agency. Under this analysis, we can still make Russo and Williamson's distinction. However, as it now becomes a question of the employment of causal relationships, rather than one dealing with their construction, we cannot seek historical support for it from the historical literature dealing with particular cases of causal inference. Instead, if we want to support this distinction we need to look for evidence in contexts where causes are used, rather than made. Doing this in detail seems beyond the scope of this thesis, so instead, I will claim here that I think this distinction should be interpreted as a claim about the use of causes, rather than one about their construction or evaluation.

Perhaps these different senses of, and reasons for seeking, causes have been a possible source of the conceptual difficulties previously experienced in interpreting causality in medicine. In fact, as Russo and Williamson identify, several recent reviews of medical causation have identified a range of meanings or senses of causation, without reaching either consensus or a conceptually satisfactory position (Russo and Williamson, 2007, 158).[2] Given both their claim about the general importance of causality in medicine (**RW1**) and the likely importance of it to a number of more specific aspects of medical practice (suggested by **RW2-3**), this failure of previous accounts seems profoundly problematic. Russo and Williamson too find this troubling:

RW4. Current accounts of causality in medicine are unsatisfactory

It is from this position that Russo and Williamson begin the constructive aspect of their account. First, perhaps some of this confusion about causality results from the many senses of 'cause' encountered in medicine. In addition to their distinction about the sorts of causes (**RW2**) and the uses to which these causes are put (**RW3**), Russo and Williamson suggest a further clarification about causes versus causal evidence. In this respect, they suggest that we should take care not to conflate the nature of causal relations themselves with their supporting or constitutive evidence:

[2] '... recent review articles try to extrapolate a concept of cause as it is used in scientific practice ... They isolate several, none of which seems to be fully satisfactory or attract a firm consensus. Parascandola and Weed (2001) isolate five different meanings of "cause": production, necessary cause, sufficient-component cause, probabilistic causation, and counterfactual. None of these can alone account for all the possible ways in which causes operate.' (Russo and Williamson, 2007, 158). See also my brief review of medical schemes of causality in section 1.2

> We shall argue that although causal relations are inferred from mixed evidence, including knowledge of mechanisms and of probabilistic relations, one ought not to confuse the causal relation itself with the types of evidence used to support putative causal relationships.
>
> (Russo and Williamson, 2007, 158)

Thus, they make the claim that causality and the evidence for causality should be distinguished:

RW5. The nature of causality is distinct from the evidence used to construct causal claims

Now this claim seems, initially, trivialy true. Of course a causal process is distinct from the evidence used to infer the existence or nature of that causal process. However, in the light of later claims (in particular **RW11**), this claim begins to make more sense. It is, essentially, directed against the claims of causal pluralists. Russo and Williamson first argue that, there exist both mechanistic and probabilistic types of evidence:

RW6. 'Evidence is constituted by two complementary elements: probabilities and mechanisms.'[3]

(Russo and Williamson, 2007, 159)

However, the double-edged notion of causal evidence (**RW6**) does not - because of **RW5** - imply that these different notions of evidence support the existence of two distinct notions of causality.[4]

So we require both probabilistic and mechanistic evidence to make causal claims (Russo and Williamson, 2007, 159; 161), but both these types of evidence support a single type of causal claim.[5] Why are both required? What functions do these types of evidence perform? First, probabilistic evidence tells us about the effects of causes via their dependencies.[6] After all, if we want to develop an understanding of disease aetiology or therapeutic strategies then our causes must make some sort of difference

[3] Note that while Russo and Williamson give various examples of mechanisms, they do not define a mechanism, nor do they relate their conception of a mechanism to any particular philosophical concept of mechanism. I will attempt to expand their account in this direction in chapters 5 and 6. They also revise their position on the role of probabilities in Russo and Williamson, 2010, by suggesting that evidence of difference-making, rather than of probabilities, is what is required to accompany evidence of mechanism. I expand upon this in sections 4.5.1 and 6.5ff.

[4] As Russo and Williamson argue: 'Suppose that the pluralist advocates two notions of cause, a mechanistic, $cause_1$, and a probabilistic, $cause_2$. Take any particular causal claim, e.g., "smoking causes cancer", that the pluralist cashes out in terms of one or other of these notions but not both... the evidence for this claim is multi-faceted, consisting of observed dependencies and mechanistic/theoretical considerations. But the pluralist's analysis of this claim will be single-faceted, say "smoking is a $cause_1$ of cancer". But then the pluralist opens herself up to the epistemological problems of monism. If this particular use of "cause" is mechanistic, $cause_1$, then how can it be that, even when the mechanism is established and uncontroversial, further probabilistic evidence is cited in support of the causal claim? However, if the use is probabilistic, $cause_2$, why are mechanisms invoked as evidence, even when there is ample probabilistic evidence? The pluralist can't explain the variety of evidence for the claim: if pluralism is right, it should be possible that the evidence just be mechanistic, or just be probabilistic.' (Russo and Williamson, 2007, 167).

[5] Russo and Williamson have recently suggested that, in fact, it is only 'typically' the case that both kinds of evidence are required (Russo and Williamson, 2010, 15)

[6] 'The probabilistic aspect is crucial because, in the health sciences, causal claims are used for prediction, diagnosis, and intervention; for these modes of inference to be possible, a cause needs to make a difference to its effects, i.e., there needs to be some appropriate probabilistic dependence.' (Russo and Williamson, 2007, 159)

in terms of their effects. We determine the presence and magnitude of this difference from probabilistic evidence. Probabilistic correlations are also heuristically important guides to the likely presence of some sort of causal relationship. In other words, correlations themselves are good - but not infallible - evidence for causation.[7]

RW7. Probabilistic evidence is required because causes need to make a difference to their effects. Correlations also suggest the possibility of a causal relationship.

The suggestion that correlations are important indicators of the existence of causal relationships is an interesting one. For instance, if we know that a particular disease increases in frequency with a particular behaviour, and reduces when that behaviour is excluded, then we have some limited causal knowledge about the relationship between the disease and the behaviour. Depending on, say, the personal impact of refraining from that behaviour, we may recommend abstention from it in order to prevent the disease without worrying too much about *how* it causes the disease.

This sort of strategy is often described as 'risk factor' epidemiology, and it can be very effective. For instance, consider the example of lung cancer and watching television. Say that an individual who watches a great deal of television is more likely to smoke cigarettes than an otherwise matched control. We would then, if we were to carry out an observational study, see higher rates of lung cancer in a group of television watchers than non-television watchers. We might therefore conduct a second, interventional, study in which television watchers were asked to decrease their level of viewing. This might well decrease the number of cigarettes they smoked, leading to a decrease in lung cancer incidence. By probabilistic evidence alone we might therefore suggest that individuals should refrain from watching television in an attempt to reduce their chance of developing lung cancer.

However, while these studies imply some kind of causal link between television and cancer upon which we can intervene, it might well be a more effective strategy to intervene directly upon cigarette smoking itself, which in this case is a more direct cause of lung cancer than television watching. In general, if we are to develop detailed causal knowledge that sustains effective, precise intervention, then we need some further evidence that tells us about such confounding variables.[8]

This correlation between television and lung cancer is also non-explanatory, in that it tells us nothing about the manner in which consuming television is supposed to cause cancer. In the light of **RW2-3**, we need some other form of evidence to make our causal claims reliable. Russo and Williamson suggest that this other form of evidence is that of mechanisms:

RW8. Mechanisms are required because they explain causal dependencies

[7] 'Although, as is well known, correlation does not prove causation, strong correlations can be good evidence for the presence of a causal relation.' (Russo and Williamson, 2007, 162).

[8] In this case, smoking is a confounding variable for the observed correlation between television consumption and the development of lung cancer. In general terms, confounders are '...factors (exposures, interventions, treatments, etc.) that explain or produce all or part of the difference between the measure of association and the measure of effect that would be obtained with a counterfactual ideal.' (Rothman et al., 2008, 58).

Evidence of mechanism tells us about causal dependence.[9] This seeks to explain the internal workings of our evidence of difference making, in a way that rationally permits us to generalise and explain our findings. For example, similar types of dependencies are likely to have similar causal mechanisms.[10] There are further methodological advantages to considering these two types of evidence. Mechanisms also do the important work of negatively constraining the permitted types of causal processes that we might consider explanatory for our probabilistic dependency. In the example above, we would begin to distrust the correlation seen between television and cancer, given the lack of a plausible mechanism linking television exposure and lung cancer:

RW9. Mechanisms are useful because they allow us to generalise a causal relation and to impose negative constraints on probabilistic causes. No plausible mechanism suggests no cause.[11]

So Russo and Williamson claim that causation in medicine depends on these two types of evidence. Probabilities play the role of showing the existence and strength of a particular cause on an effect, while mechanisms demonstrate the manner in which these probabilistic correlations function. With this basic framework in place, Russo and Williamson turn to the sort of causality that this evidential pluralism supports. First, they argue against conventional, monistic accounts of causality,[12] on the grounds that all of them fail to accommodate the sorts of evidential pluralism that they, and medical practice, demand.

RW10. Traditional causal monism cannot accommodate this evidential pluralism.

Broadly, mechanistic accounts of causality cannot explain the reliance on difference-making evidence that appears to be so important in defending causal claims in medicine. That is, the role played by finding and evaluating correlations that plays such an historically important role cannot be accounted for by mechanistic monists. Similarly, difference-making accounts cannot account for the use of mechanistic evidence in practice. Thus, Russo and Williamson argue, any monistic interpretation of causality in medicine that depends upon any one of these accounts appears false. They also argue that any simple admixture of both mechanistic and difference-making accounts is false too. It is simply not the case, argue Russo and Williamson, that a number of different types of causality are a feature of the medical literature:

---

[9] '...the mechanistic aspect is required because mechanisms explain the dependencies, and in the health sciences causal relationships are also meant to be explanatory.' (Russo and Williamson, 2007, 159)

[10] Take, for example, the research programme that developed around HSV as a cause of cervical cancer. While this was eventually demonstrated to be mistaken, it was in part driven by the heuristic that many other herpesviruses appeared to play a causal role in the development of tumours

[11] '...mechanisms allow us to generalise a causal relation: while an appropriate dependence in the sample data can warrant a causal claim "C causes E in the sample population", a plausible mechanism or theoretical connection is required to warrant the more general claim "C causes E". Conversely, mechanisms also impose negative constraints: if there is no plausible mechanism from C to E, then any correlation is likely to be spurious.' (Russo and Williamson, 2007, 159)

[12] Russo and Williamson include both mechanistic and difference-making accounts of causality here. They mention, for the class of monistic mechanistic theories, the conserved quantities (CQ) theory of causality (Salmon, 1998; Dowe, 2000). Monistic difference-making theories are represented by a range of probabilistic theories (e.g. Suppes, 1970), counterfactual theories (e.g. Lewis, 2000) and agency theories (e.g. Price, 1991).

> ... only a single notion of cause is used in the health sciences. More specifically, health scientists use two types of evidence for a single causal claim, '*C causes E*', not for two different types of causal claim, *C-causes*$_1$*-E* and *C-causes*$_2$*-E*. Because only one notion of cause is used in the health sciences, pluralism is false.
>
> (Russo and Williamson, 2007, 165)

This gives us Russo and Williamson's 11th claim, arguing against any form of ontological or conceptual causal pluralism:

RW11. Medicine is not pluralistic about causality

This refutation of causal pluralism gains support from two sources. Firstly, the homogeneity of causal language in the medical literature seems to support causal monism (Russo and Williamson, 2007, 166).[13] Second, they suggest that pluralistic accounts will inherit the difficulties of monistic accounts in accomodating alternative forms of evidence (Russo and Williamson, 2007, 166-7). Thus Russo and Williamson appear stuck. They need a monistic interpretation of causality that can accommodate a plurality of evidence. Their solution is to turn to an epistemic theory of causality:

RW12. '... an epistemic account of causality is required to capture the full complexity of causal evidence' (Russo and Williamson, 2007, 169)

Russo and Williamson suggest that an epistemic theory of causality, in which causal relationships are identified as '... the causal beliefs of an omniscient rational agent' (Russo and Williamson, 2007, 168),[14] is the correct interpretation of causality in this situation. This epistemic theory means causality should be understood as arising from causal epistemology:

> The epistemic theory understands causal relationships in terms of rational beliefs as follows. First, it gives primacy to an account of how an agent's evidence determines which causal beliefs she should adopt. Second it analyses the causal relation to be the set of causal beliefs that an agent with total evidence should adopt. Thus causality itself is determined by causal epistemology... How, then, does evidence constrain rational causal beliefs? Roughly speaking, an agent's causal belief state should satisfy the following conditions: (i) her causal beliefs should account for all known dependencies that are not already accounted for by known non-causal relationships; (ii) her causal beliefs should be compatible with her other

[13] Referring to a sample passage from the recent medical literature, Russo and Williamson say: 'Here causal language is used ("the same concentration in the target tissue may cause the same level of effect in rats and humans"); the scientists are clearly looking for mechanisms ... and for difference-making, too ... Likewise, when Kuhl (2006, 680) says "About 10% of breast cancers are 'hereditary', i.e., caused by a pathogenic mutation in one of the 'breast and ovarian cancer susceptibility genes' (BRCA)" the word "caused" is associated both with a genetic mechanism and with making a difference to 10% of breast cancer patients.' (Russo and Williamson, 2007, 166)

[14] This quotation continues: 'This gives a view of causality that is analogous to the objective Bayesian view of probability, according to which probabilistic beliefs are determined by an agent's evidence, and probabilities themselves are just the beliefs that an omniscient agent should adopt.' (Russo and Williamson, 2007, 168)

> knowledge (including knowledge of mechanisms); and (iii) she should not have any causal beliefs that are not warranted by her evidence.
>
> (Russo and Williamson, 2007, 167-8)

This (monistic) theory of causality can also account for the evidential pluralism that Russo and Williamson claim:

> The epistemic theory of causality can account for this multi-faceted epistemology, since it deems the relationship between the various types of evidence and the ensuing causal claims to be constitutive of causality itself.
>
> (Russo and Williamson, 2007, 168)

Like Russo and Williamson, I will concentrate on the epistemic aspects of this relationship between causality and evidence. However, unlike Russo and Williamson, I will not give an account of a general epistemic theory of causality. In other words, I will neglect the implications of **RW12**, and will instead concentrate on the kinds of epistemic devices that allow us to conceptualise the relationship between evidence and causality. Thus the key feature of my account is the same as Russo and Williamson's. This is the distinction between causality and evidence (**RW5**). While, they suggest, there is only one sort of cause in medicine, this single, monistic causation is not supported by just one sort of evidence. As they note, accounts of causation based upon a single type of evidence fail. Therefore, the decision to accept or reject a monistic causal claim requires the consideration of a plurality of evidence.

However, here there is a difficulty that must be addressed: current medical practice does not seem to be pluralistic about evidence. While Russo and Williamson argue in support of evidential pluralism by highlighting an apparent pluralism in the kinds of evidence used to support causal claims (see footnote 13, page 83), this claim is compromised by the widespread use and acceptance of Evidence-Based Medicine (EBM).

EBM is defined as 'the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients'(Sackett et al., 1996). In this context, 'current best evidence' has a specific meaning. According to proponents of EBM, it is possible to rank evidence about the efficacy of interventions in terms of its reliability. This is achieved by using so-called hierarchies of evidence. These systems of ranking provide healthcare workers with 'a measure of the trust that can be placed in the recommendations' (Evans, 2003, 79) produced by clinical studies. In fact, for the EBM practitioner, the best kind of evidence is usually that provided by either a randomised control trial (RCT) or - even better - a meta-analysis of RCTs (Guyatt et al., 2008):

> Because the randomised trial, and especially the systematic review of several randomised trials, is so much more likely to inform us and so much less likely to mislead us, it has become the "gold standard" for judging whether a treatment does more good than harm.
>
> (Sackett et al., 1996)

While it is true that there are a great many different evidence hierarchies without exact consensus about precisely how such evidence should be ranked, in general RCTs and meta-analyses are regarded as the most reliable form of evidence. This position therefore lends strong support to a claim that medicine is monistic, rather than pluralistic, about evidence. How can this difficulty for the RWT be overcome?

First, EBM is generally not presented as an explicitly causal methodology. So a lack of evidential pluralism here does not necessarily present difficulties for the proponent of the RWT. Taking this tack as a way of defending the RWT would be disingenuous, though: while EBM does not purport to be a causal methodology, as it looks to investigate the relationship between an intervention (a cause) and an outcome (effect) it is really concerned with causation (Bluhm, 2005). So this line of argument is no adequate defence against this criticism.

However, closer consideration of the hierarchies of evidence employed in EBM practice suggests a more adequate response. In fact, I argue that the EBM hierarchy of evidence is not an hierarchy of evidence at all. I (and others - Bluhm (2005), for example) claim that the hierarchy of evidence is really a hierarchy of experimental methodologies, rather than of evidence itself. Studies are ranked by means of reference to the way they were carried out, rather than the types of thing they provide evidence about. The only evidence worth having, argues the EBM-proponent, is that which arises from an RCT, or meta-analysis. It is therefore monism about methodologies that is promoted by the EBM movement, not monism about evidence.

As these two distinctions between types of evidence are likely to cross-cut each other in complex ways, this suggests that the evidential pluralism required by Russo and Williamson might be somewhat compatible with the methodological monism proposed as a part of EBM. In fact, as Russo and Williamson require pluralism about the kinds of things we have evidence of, rather than the kinds of methodologies employed to produce this evidence,[15] I will argue that the type of methodological monism required by EBM conceals a profound pluralism about evidence itself.

Take, for example, a particular RCT. It is not the case that this trial will be simply concerned with one kind of evidence, either mechanistic or probabilistic. While investigators will report the results of this trial in terms of statistical data about the relationship of cause and effect, and this technique might perhaps lead us to the conclusion that RCTs are therefore concerned only with probabilistic evidence, a closer investigation of these results convincingly demonstrates the central role of mechanisms in producing and evaluating this evidence.

How does mechanism participate in the construction of an RCT? First, the kinds of correlations that RCTs aim to detect are generally not simple. Instead, trials are usually designed to find out what happens, in terms of change in outcomes and intermediate markers of outcomes, what happens when an intervention is performed. For example, (O'Shaughnessy et al., 2011) - an entirely random example from a recent medical journal - looked to demonstrate the differing effects of using iniparib ('a small molecule with PARP-inhibitory activity' (O'Shaughnessy et al., 2011, 205)) as an adjunct chemotherapeutic agent for treating breast cancer. In this paper, patients were selected for inclusion on a range of mechanistic

[15] I develop a similar distinction in section 6.4, along lines first developed by McKay Illari (McKay Illari, 2010). I will leave the details of this distinction for that section.

features of breast cancer (so they could only participate in the trial if they had 'triple-negative' breast cancer),[16] as well as a number of general mechanistic features of fitness (adequate renal and hepatic function, for example). These factors were selected because - mechanistically - it is the case that presence of these particular things changes outcome. Thus recruitment to this trial was dependent on mechanistic reasoning. Similarly, the complex set of outcome measures collected (clinical benefit, rate of overall response, progression-free survival and median overall survival) are themselves causally complicated. If we deny the role of causal mechanisms in informing the design of such a clinical trial, we cannot account for the particular choice to observe or control these factors. Indeed, the choice of intervention in this case was itself dependent on a mechanistic understanding of the actions of iniparib. As no other trials of this agent had previously been performed, what other kind of knowledge could guide the decision to see if this molecule - and not some other - might improve cancer survival? As the authors suggest, it is because this agent was known to interfere with Poly-ADP ribose polymerase (PARP) - a protein that controls cell death - that made it a suitable candidate for such a study.

Similarly, interpreting these results requires some mechanistic knowledge about breast cancer and breast cancer interventions. It is not simply the case that this agent will lead to likely improvements in survival for all individuals with breast cancer. Instead, it is only a certain subgroup - those with triple-negative breast cancer - that this study provides evidence of benefit for.

In short, RCTs are monistic about methodologies, but not about evidence itself. To suggest that EBM methodology does not require any consideration of mechanism is to miss the point entirely. Without evidence of mechanisms, the construction and evaluation of an RCT would be impossible.

Finally, Russo and Williamson's account is really concerned with the acceptance, rather than the construction, of causal claims.[17] However, given the account of causal discovery given in 1.3, I will develop the RWT such that it can give an account of some parts of discovery. I will exclude from this treatment only the very earliest stage of discovery - that of suspicion. As I will discuss in section 6.3.1, the treatment of evidence at this early stage is distinctly unlike that found at other points. So I want to specifically exclude just the very earliest parts of discovery from the RWT.

Now, though, I will turn to some philosophical objections to Russo and Williamson's arguments.

4.3 Philosophical counterarguments to the RWT

There are two main types of philosophical objection to the RWT. First, some authors argue that Russo and Williamson are mistaken to require mechanisms, or mechanistic evidence, as a necessary component of causal claims. Here, I will review two such objections by Broadbent (2010) and Howick (2010).[18] Second, Weber (2009) has criticised **RW10**, the claim that evidential pluralism cannot be accounted for

[16] i.e. that the tumour cells expressed neither oestrogen receptors, progesterone receptors nor human epidermal growth factor 2 receptors

[17] '...two different types of evidence - probabilistic and mechanistic - are at stake when deciding whether or not to accept a causal claim.' (Russo and Williamson, 2007, 163)

[18] Gillies, while broadly supporting the RWT, has suggested that a plausible mechanism may well be sufficient warrant to support RWT-type causal claims (Gillies, 2010). I think this is an unnecessarily weak version of the thesis, and one that may well lead to real operational difficulties - as the HSV case suggests. I will go on to discuss plausible mechanisms in section 6.4.1.

by traditional monistic theories of causality.

Broadbent's criticism of the RWT depends upon the interpretation of it as either a descriptive or a normative account (Broadbent, 2010).[19] First, if the RWT is interpreted as a descriptive theory, then (as it stands), it does not support any claims about causation or causal inference:

> ...a purely descriptive reading of Russo and Williamson's claim renders it largely irrelevant from a methodological point of view, and does not justify or explain why they themselves treat it a motivation to seek a theory of causation, apt for the health sciences.
>
> (Broadbent, 2010)

Thus the descriptive RWT is irrelevant. However, Broadbent's suggestion about a normative reading is even more damning. He argues that, if the RWT is interpreted normatively, then the requirement for mechanistic evidence is faulty (i.e. against **RW6** and **RW8**). He provides both a methodological and an historical argument to support this claim. His methodological argument is as follows:

> Modern epidemiologists set very high store in some methods, such as the randomized control trial, which involve no requirement to identify underlying mechanisms.
>
> (Broadbent, 2010)

While historically, Broadbent refutes the normative interpretation of the RWT by reference to the Semmelweis case:

> ...the lesson Russo and Williamson draw from the Semmelweis episode is that Semmelweis's theories ought not to have been accepted until knowledge of underlying mechanisms was obtained. (This explains why they offer a theory of causation intended to justify this stance...then they are committed to the startling view that, had germ theory not come along and the underlying mechanism remained a mystery, we today would be rational to dismiss Semmelweis's work, no matter how much evidence we had gathered in the meantime about the efficacy of disinfecting hands.)
>
> (Broadbent, 2010)

He also suggests the two arguments, methodological and historical, seem to complement one another:

> In particular, the evidence for the efficacy of his [Semmelweis's] proposed intervention - disinfecting (not merely washing) hands - is extremely strong. And replication would have made it stronger, without necessarily advancing knowledge of underlying mechanisms.
>
> (Broadbent, 2010)

[19] 'I am unsure whether it is meant as a descriptive account of causal inference (then and now), or as a normative account of the standards which ought to be used when deciding whether to infer causation.' (Broadbent, 2010). I support a normative reading of the RWT, as I discuss in section 7.3.

I think this criticism is based on a mis-reading of what mechanistic evidence is. First, his methodological argument seems to badly neglect the role of mechanistic evidence in informing the methodologies of epidemiological investigation. As I will discuss in sections 4.5.2 and 6.5, the design of detailed epidemiological studies critically depends upon mechanistic evidence. Otherwise, the manner in which particular factors are selected for study in clinical studies seems nothing short of miraculous. It seems highly implausible that drugs, vaccines, immunological tests and so on arise in the absence of any mechanistic understanding whatsoever.

Now Broadbent's historical argument seems similarly flawed. He offers the following counterfactual: 'had germ theory not come along and the underlying mechanism remained a mystery, we today would be rational to dismiss Semmelweis's work, no matter how much evidence we had gathered in the meantime about the efficacy of disinfecting hands.'

I disagree with this counterfactual statement for two reasons. First, assuming that the germ theory had not arisen, it would be rational to argue that we did not understand the manner in which disinfecting hands caused a reduction in the incidence of puerperal fever, not that this intervention should not be performed. Failing to establish a causal claim does not imply that a particular research programme should be ignored or rejected.

Second, the conditions necessary to instantiate this counterfactual appear highly implausible. For the germ theory to not be developed, yet for medical practice to proceed in such a way as to produce much evidence showing the benefits of disinfecting hands, appears to me very unlikely. This is because the very concept of disinfection itself is unintelligible without some theoretical understanding of what is happening in terms of germs. Without the germ theory,[20] the only possible means of discriminating between these two behaviours would be a contingent, methodological one, such as a classification in terms of the cleansing agent employed. I suggest that this would render the Semmelweisian methodology exquisitely fragile, and thus unlikely to survive (i.e. produce similar benefits in survival) any translation to different practical contexts. In this case, unless Semmelweis's methodology was followed exactly, similar improvements would not accrue. I thus reject the assertion that a programme of disinfection, capable of producing similar improvements in clinical outcomes, could have come about in the absence of some kind of germ theory of disease.

Howick objects to the RWT on similar grounds to Broadbent. His general argument is in support of an increased role for mechanistic evidence in EBM practice.[21] However, this support for mechanisms is strongly tempered by the limitations of mechanisms as causal evidence when compared to evidence arising from comparative clinical effectiveness studies, such as RCTs. While Howick thinks that mechanisms may be useful supporting evidence for causation, they should not be used to conduct causal inference in isolation, neither should they be relied upon to overturn alternative, superior, forms of evidence, nor are they a necessary part of causal inference.

First, he argues against the exclusive use of mechanisms in formulating causal claims. Howick first

[20] Or some other mechanism capable of explaining, in general terms, the difference between washing and disinfecting hands.

[21] '...a central thesis of this paper is that high quality mechanistic reasoning should be allowed a more prominent place in Evidence-Based Medicine' (Howick, 2010).

defines just what it means to reason in this way:

> **Mechanistic reasoning**: involves an inference from mechanisms to claims that an intervention produces a patient-relevant outcome. Such reasoning will involve an inferential chain linking the intervention (such as antiarrhythmic drugs) with a clinical outcome (such as mortality).
>
> (Howick, 2010)

He then goes on to outline just how this kind of causal reasoning is likely to lead to faulty causal conclusions:[22]

> ...mechanisms, by themselves, do not count as evidence. Accounts of how the heart, brain, and liver work will not tell us what will happen to these mechanisms under intervention. To count as evidence for claims that interventions have their putative patient relevant effects, we must infer from (supposed) knowledge of the relevant mechanisms to claims that treatments have patient relevant effects. In order for this inference to be acceptable, we must know what happens to each of the relevant mechanisms under intervention.
>
> (Howick, 2010)

Second, he then argues against treating mechanistic evidence as a superior form of evidence to that generated from comparative effectiveness studies, such as RCTs. In situations where evidence arising from comparative clinical effectiveness studies and mechanistic evidence are contradictory, Howick argues against the general overturning of evidence of comparative clinical effectiveness by mechanistic evidence:

> ...there are reasons to be cautious about using mechanistic reasoning to rule out or reduce our confidence in results from comparative clinical studies, and moreover the principle of total evidence does not imply that we require mechanistic reasoning to rule out implausible hypotheses.
>
> (Howick, 2010)

Finally, he argues that it is not the case that mechanistic evidence is necessary to support causal claims. As this contradicts **RW8** - and therefore compromises the RWT - it is this argument that I will focus the remainder of this section on.[23] Howick provides three arguments in support of this claim - two directed against the historical component of the RWT, and one against the theoretical.

Howick presents two arguments against the historical component of the RWT, which suggest that it is unwise to require mechanistic evidence in the presence of high-quality comparative clinical research

[22] I agree that you cannot understand disease aetiology or therapeutic interventions by describing physiology. I therefore develop a very similar position in further chapters.

[23] Howick's two other arguments seem similar to those made by Russo and Williamson. As per **RW10**, Russo and Williamson too seem to concur that mechanisms alone are an inadequate means of assessing causality. Nor do they suggest that one form of evidence should generally overturn another.

for the following reasons. First, that requiring mechanistic evidence may delay the clinical introduction of effective techniques (Semmelweis, *Helicobacter pylori*), and second, because a wide range of aetiological and therapeutic causal hypotheses have been accepted and successful without any mechanistic content.[24] Because of these two types of empirical argument, Howick suggests that the mechanistic component of the RWT is redundant.

His theoretical argument is as follows: Russo and Williamson claim that mechanisms avoid confounding, and thus rule out spurious relationships. But, according to Howick, this is not true. In fact, he suggests that comparative clinical effectiveness studies are better at ruling out spurious relationships than mechanistic reasoning. He provides an impressive list of historical cases where clinical research has overturned incorrect mechanistic reasoning. But this claim appears to support the RWT: Russo and Williamson absolutely do not claim that causality is established through mechanistic reasoning alone. Instead, both probabilistic and mechanistic evidence are required. Giving examples where evidence of one form overturns evidence of another is not an effective argument against their thesis. Howick then suggests that looking for both evidence of mechanism and evidence of clinical effectiveness might lead to better causal claims:

> One might alter the claim that mechanistic reasoning is required to rule out spurious relationships and argue instead that hypotheses that are supported by both comparative clinical studies and mechanistic reasoning are less likely to be spurious than hypotheses supported by one type of evidence alone. Such a claim might rest on the plausible premise that since each type of evidence suffers from different potential pitfalls, a hypothesis that is supported by both types of evidence is less likely to have been confounded. This weaker claim is altogether acceptable and indeed underwritten by the principle of total evidence. However, in its altered form it can no longer be interpreted as Russo and Williamson's thesis.
>
> (Howick, 2010)

In summary, while Russo and Williamson insist on both mechanistic and probabilistic evidence,[25] Howick suggests that mechanistic evidence is just a useful adjunct to understanding causality, rather than an essential component. In fact, as his historical cases suggest, requiring mechanistic evidence may sometimes not only be redundant, but may actually cause harm.

A suggestion for this difference of opinion might be as follows: Howick disagrees with Russo and Williamson because they understand mechanistic evidence differently. Howick takes a narrower view of what mechanistic evidence is than do Russo and Williamson. As per his definition of **mechanistic reasoning** above, Howick suggests that mechanistic evidence is concerned with inferring the relationship that obtains between an intervention and its clinical effects. Russo and Williamson, on the other hand,

[24] These include Pott's discovery of the causal relationship between exposure to soot and cancer of the scrotum in chimney sweeps, Jenner's discovery of vaccines against smallpox, Snow's epidemiological investigations of cholera, Finlay's discovery that yellow fever could be prevented by killing mosquitoes, the discoveries of aspirin, anaesthesia and steroids and the (more recent) discovery that electrical deep brain stimulation improves the symptoms of various movement dystonias.

[25] Although it seems that Russo and Williamson may now agree with Howick on this point, as they have subsequently modified their argument such that claim 'typically' both kinds of evidence are required (Russo and Williamson, 2010, 15).

seem to take a much broader view of what constitutes appropriate mechanistic evidence (see footnote 13, page 83.). Under the RWT, mechanistic evidence is only part of such an inference, rather than wholly constituting it. In fact, Russo and Williamson require that additional, probabilistic evidence - such as might result from performing an RCT - will also be required to make such an inference. I support a similarly broad reading of what constitutes mechanistic evidence. Mechanistic evidence is employed extensively in the design of clinical trials, their interpretation and the implementation of their results. For example, when we attempt to discriminate high- from low-quality comparative effectiveness research, we rely on employing a great deal of mechanistic evidence to tell us about the appropriate use of 'employed placebo controls, double blinding, allocation concealment and also checked for baseline differences between comparison groups, and hence avoided known plausible confounders' (Howick, 2010). Designing a study such that these virtues are sought requires mechanistic knowledge about the particular case. Otherwise, the choice of an appropriate placebo, of blinding procedures, of baseline comparisons and the avoidance of confounders would appear miraculous.

Weber criticises the RWT by attempting to demonstrate that Russo and Williamson's claim that evidential pluralism cannot be accommodated by traditional causal monism (**RW10**) is incorrect. He suggests that Giere's theory of probabilistic causality can, in fact, accommodate both mechanistic and probabilistic evidence when it comes to establishing causal claims in scientific practice:

> I have presented Giere's theory of probabilistic causation, and argued that it can account for the use of mechanistic evidence in various contexts in which scientists use such evidence. This undermines the thesis of Russo and Williamson that probabilistic theories of causality cannot account for the use of mechanistic evidence.
>
> (Weber, 2009, 293)

Giere's basic idea is this. Say we wish to establish a causal claim linking a dichotomous causal variable, C, with an effect, E, in a population U. We should do this by comparing the probability of E arising in two hypothetical populations, X and K, where all members of X do C, and all members of K do $\neg C$. C will then be a *positive causal factor* for E in U iff $P_X(E) > P_K(E)$, a *negative causal factor* for E in U iff $P_X(E) < P_K(E)$, and *causally irrelevant* iff $P_X(E) = P_K(E)$.

Weber defends this claim through examination of two different situations in causal evaluation. First, he discusses the case where we wish to establish causal claims in the absence of probabilistic evidence (Weber, 2009, 285-6). As, in this case, our Gierean claims would depend upon thought experiments, Weber claims that: '...to make a Gierean causal claim about a population, it is perfectly rational to gather mechanistic evidence, since that evidence can help...to establish the claim' (Weber, 2009, 286). This, as Weber admits, seems to be perfectly compatible with the RWT.

His next situation is where both mechanistic and probabilistic evidence are present, and we wish to establish a causal claim. Weber claims that the Gierean theory can also account for the use of mechanistic evidence within a probabilistic framework (Weber, 2009, 287ff). His argument is as follows. Ideally, when we wish to establish a Gierean causal claim, we would perform randomised trials on our

population of interest. However, there are various operational and ethical strictures on actually following this methodology in practice. So there will be many occasions where an ideal, randomised trial is impossible. In situations like these, we must rely on observational studies on our population of interest, or on randomised trials on a different population. Weber argues that Giere's theory can account for the use of mechanisms that intervene between the actual results of these trials, and the human population of interest, in both these situations. That is, mechanisms can account for confounding variables (in the observational trial case) and also deal with questions of external validity (in the case of trials on other populations). As he suggests:

> ...for a scientist who wants to make Gierean causal claims about populations, it is perfectly rational to look for mechanistic evidence: it will be useful for processing the results of prospective and retrospective studies.
>
> (Weber, 2009, 288)

and:

> a scientist who wants to make a Gierean causal claim, needs a warrant to link the animal data to the hypothetical human populations. Mechanistic evidence can provide this link. So it is rational for such a scientist to look for mechanistic evidence.
>
> (Weber, 2009, 289)

Now these arguments certainly do seem to be a good representation of scientific practice. But these claims also seem compatible with the RWT. Weber does not make the case that the Gierean probabilistic causality fully accounts for this mechanistic evidence. Rather, this seems to be an argument in support of employing Giere's theory to understand probabilistic evidence as part of the RWT. For, while it may be rational for a Gierean to employ mechanistic evidence to support their evaluation of probabilities, this process is not part of establishing an exclusively Gierean causal claim. Quite the opposite: a high-quality Gierean causal claim (one that, say, employs a randomised trial on the actual population of interest) cannot account for the use of mechanistic evidence at all.[26]

So I do not think this objection seriously damages the RWT. While Weber makes a convincing case that it may well be rational to employ both mechanistic and Gierean methodologies in establishing a good causal claim, he does not adequately make the case that the Gierean theory itself can account for the use of mechanistic evidence. As Weber suggests that it is only Giere's account that has this advantage, and no other theory of probabilistic causality, this does not seem a general objection to the RWT.

4.4 The RWT as an empirical proposition

While Russo and Williamson make their claims in a theoretical manner, their one-cause, plurality-of-evidence view gains general support from the medical literature on causation (c.f. section 1.2). For

[26] Williamson has argued similarly against Weber's claim. He also suggests that most of Weber's examples that feature mechanistic evidence in support of Gierean causal claim are rather weak, in terms of their probabilistic evidence. Williamson, paper to the Second UCL-Kent workshop on causality, 27th January 2010.

instance, both the HKP (see, for instance, Carter, 2003) and the Hill criteria (Hill, 1965) require a range of evidence, yet do not distinguish between types of causation based upon these differing evidential inputs. They gain further empirical support for their thesis from the position of mechanisms and probabilities. In fact, Russo and Williamson argue, these types of evidence are so important, and so widely used, as to be a commonplace in the medical literature.

At the outset, I think that the empirical validity of the RWT can be demonstrated in both cases of viral oncogenesis detailed above. However, I will for the time being only discuss the EBV case, while ignoring the history of cervical cancer. The reason for this partial reading of the history is straightforward. As I have already mentioned, the cervical cancer case is complicated by the HSV episode. Thus, as it stands, this case seems to present a counterexample to Russo and Williamson. I think this impression is misleading. But in order to demonstrate exactly why this is so, I need first to do some minor elucidation of the RWT, which will follow, along with an analysis of the HSV case, in section 4.5. For now (and for the sake of clarity) I will focus on the EBV case.

An important stage in the discovery of the causal relationship between EBV and BL was the process of mechanism construction. This process clearly demonstrates the fundamental evidential pluralism required to make good mechanisms. In it, we see the exploitation of a wide variety of experimental and observational strategies. Briefly, and for the sake more of illustration than completeness, we see mechanistic features arising from clinical observations on cases of the tumour syndrome, from histopathological investigations of these same cases, from interventions upon tumour cells (notably the establishment of cultured lines of BL cells) and from a number of specific virological techniques (including observations of viral CPE, attempts to perform viral cultures on material derived from BL tumours, and the use of numerous immunological technologies). Importantly, though, significant advances in mechanism construction were also made by investigators concentrating on diseases other than BL. Two unusual themes are apparent. First is the exploitation of other, possibly aetiologically related or clinically similar diseases as sources of evidence for the EBV-BL causal claim. These areas of inquiry yielded both negative constraints on the shape of the EBV-BL causal mechanism (by, say, differentiating the pathogenic mechanism of leukaemia from that of BL) as well as suggesting possible areas where the EBV-BL mechanism might be expanded (as, for instance, occurred in the work arising from the geographical correlation noted between BL and various arthropod-borne diseases). Second, this case demonstrates a vital role for the related concepts of analogy and plausibility in biomedical research. I will address the general aspects of these phenomena in later chapters. For now then, I will make a few remarks on the specifics of the EBV case.

4.4.1 The EBV mechanism

There were three problems that persistently complicated causal claims implicating EBV infection in the development of BL, and it was not until they were overcome (or at least became disregarded) that causal claims about EBV causing BL became generally accepted. The first of these was the **range of viruses** problem. In addition to EBV, many other infectious agents, generally regarded as plausible causes of cancer, were found in BL tissue. These included arboviruses, Echo 11, HSV and a range of

other agents.[27] Why should EBV be regarded as the cause of the tumour, and not these others?

The ubiquity problem. EBV was soon found to have a ubiquitous, worldwide distribution. Almost all adults were found to have serological evidence of prior infection. Yet BL appeared to have a remarkably clearly demarcated distribution. How could a virus everyone had cause clinically apparent disease in such a specific population?

The malaria problem. Geographical evidence highlighted the similarity between the distribution of certain arthropod-borne diseases and that of BL. Plausibly, this might happen in three ways. One, the causal agent of BL might be vectored by an arthropod. Two, an arthropod-borne agent such as malaria or African trypanosomiasis might be a cofactor in the development of the disease. Finally, the insect distribution might merely geographically correlate with some other causally significant environmental or climate-dependant factor. While EBV was shown not to be arthropod-borne (thus ruling out insect-transitted EBV as a solution to this problem), no clear evidence was forthcoming that enabled researchers to select between outcomes two and three.

Together, these three problems represented a substantial threat to the soundness of the developing EBV-BL mechanism. Their resolution, however, did not come about by the result of any specific mechanistic inquiry directed at them. Instead, the satisfactory resolution of these problems arose by primarily statistical means; through interpretation of the results of the 1978 prospective study (de-Thé et al., 1978).

First, the **range of viruses** problem. While researchers did not perform serology for all the possible candidate agents previously recovered from BL tissue, they did perform matched-control testing for a range of EBV antigens, for CMV, HSV and measles virus (de-Thé et al., 1978, 758-60). These results indicated that the abnormal immunological response to EBV in those later to develop BL was EBV-specific. Importantly for the range of viruses problem, the specificity of this response strongly counted against an aetiologically significant role for other viral agents in the development of BL. It further suggested interesting features of the primary EBV infection that were in need of mechanistic decomposition.

This specificity of response neutered the **ubiquity problem**. Not only was the demonstrated immune response found to be solely one of EBV, but it was specific to just one viral antigen. Individuals who later developed BL showed an increase in just one anti-EBV antibody (EBV viral capsid antigen (VCA)), and not, in general, increases in any of the other antibodies. This suggested that the primary EBV infection in those later to develop the tumour was not simply a prolonged, or especially severe form of normal EBV infection. Instead, the causally responsible primary infection was of an unusual type. This, when combined with the rather different serological profile of IM, suggested avenues of further mechanistic inquiry. Of particular importance was the way in which the infection, and thus the tumour, developed over time. It seemed that the abnormal serological response seen was likely to indicate the effects of some sort of co-factor, modulating the immune response to primary EBV infection.

Finally, the growing understanding of the temporal sequence of events in BL pathogenesis provided the means to understand the **malaria problem**. First, BL patients did not have higher levels of

[27] see section 2.4.2.4

malaria parasites at recruitment than controls (de-Thé et al., 1978, 60). However, the specific nature of the increase in VCA antibodies seen in BL indicated that BL was not the manifestation of some unknown underlying immune dysfunction. Instead, this abnormal serological response to BL suggested that malaria was acting as the final triggering event for BL in individuals with evidence of abnormal primary EBV infection (de-Thé et al., 1978, 761). Malaria therefore becomes a target for further mechanistic investigation, as suggested by attempts to prevent BL by rigorous attempts at paediatric malaria prevention.

What role did mechanism play in this *eperimentum crucis*? As we have seen, the construction and interpretation of sophisticated population-level studies critically depends on knowledge of mechanisms. This case is no exception; without the EBV-BL mechanism, this experiment could neither have been designed, nor understood, in the way that it was. As a consequence, a study not based on this kind of knowledge could not have convincingly demonstrated that EBV really was causally implicated in the development of BL. This is because, importantly, it was not the case that there was any kind of simple correlation between EBV infection and BL. Instead, EBV behaved in a completely different way in those later to develop BL compared to those who did not, in that the specific serological markers of infection differed qualitatively and quantitatively between the two groups. While it seems perfectly plausible that we could discover a simple correlation between these two things without needing any knowledge of mechanism, it does not seem plausible that we could have detected the subtle relationship that actually obtains in this case. For without knowledge of how EBV mechanistically behaved in individuals likely to develop BL and in the unlikely majority, these differences would have been neither detected nor understood. It was this detailed knowledge, linking both mechanistic and probabilistic evidence, that provided the solutions to the three problems affecting the BL-EBV mechanism, and lead to the general acceptance of EBV as the first known human oncovirus.

4.5 Historical counterexamples to the RWT

4.5.1 Causation Without Statistics - McArdle Disease

McArdle disease is a rare[28] autosomally recessive condition of abnormal muscle glycogen metabolism. Individuals with the condition suffer excessive tiredness and severe cramps on exertion. Skeletal muscle, in common with other metabolically active tissue, must be provided with a constant supply of energy. One of these sources of energy is glucose, which circulates in the blood. At rest, the supply of glucose from the blood is more than sufficient to fulfil the energy requirements of the muscle. However, during exercise the total energy demand of the muscle exceeds the maximum possible rate of uptake of glucose from the blood. Hence, the muscle must obtain glucose from another source. This source is glycogen, a large, branched polymer of glucose. Stores within normal muscle are enzymatically cleaved into glucose during times of intense metabolic demand. The muscle can then replenish its glycogen stores from

[28] Also known as myophosphorylase deficiency (OMIM 232600). It is *very* rare. While no precise prevalence figures are available, a recent study in Canada identified a rate of 1.4/100 000 births for *all* glycogen storage disorders, with McArdle disease as just one of them (Applegarth et al., 2000, e3). 55 individuals with McArdle disease are currently known in the UK (Hilton-Jones, 2001, 122), suggesting a rate of about 1/1 000 000 births. This is supported by Chen's estimate of 1/800 000 (Chen, 2001, 1530).

circulating glucose when at rest.

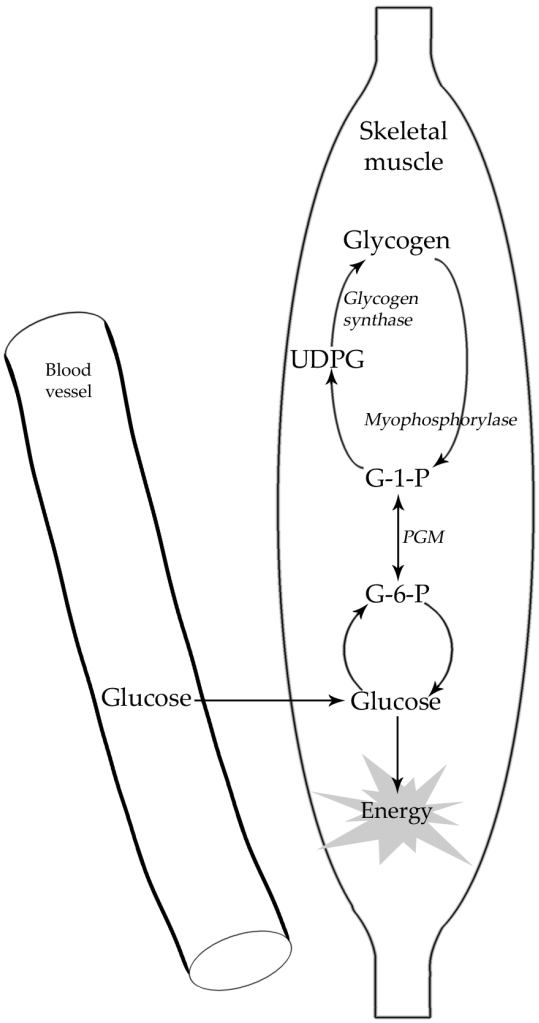

Figure 4.1: Simplified mechanism of glucose-glycogen interconversion. Note that the details of both the conversion of glucose to energy, and the conversion of glucose to glycogen have been omitted for the sake of clarity. G-1-P: glucose-1-phosphate, G-6-P: glucose-6-phosphate, PGM: phosphoglycomutase and UDPG: uridine diphosphate glucose.

A simplified pathway for the synthesis and breakdown of glycogen is shown on page 96. As can be seen, the anabolic and catabolic pathways are asymmetric. Two different enzymes exist, one to add glucose molecules to glycogen,[29] and one to remove them. McArdle disease is caused by a functional lack of myophosphorylase,[30] the enzyme which removes glucose units from the glycogen molecule. Thus, patients with the disorder can form normal glycogen, but cannot convert it back to glucose.[31] Thus we may begin to explain the brief clinical scenario above. Skeletal muscles function normally until called

[29] Glycogen synthase

[30] IntEnz Nomenclature EC 2.4.1.1. The official systematic name is 1,4-a-D-glucan:phosphate a-D-glucosyltransferase

[31] Note that the defect is confined to skeletal muscle, as the phosphorylase enzymes occurring in other tissues are coded for by different genes. The gene encoding myophosphorylase is found on the long arm of chromosome 11 (*11q13*) whereas the gene for the liver form is found on chromosome 14 (*14q21-q22*) and the brain form is found on chromosome 20 (*20p11*). A large number of mutations of myophosphorylase are known, currently at least 14 (Chen, 2001, 1538).

to strenuous activity. When this occurs, they rapidly run out of energy, leading to fatigue and pain. With the exception of these limitations to physical exercise (and their obvious limitations upon the patients activities of daily living), McArdle disease appears to be relatively benign and affected individuals have a normal life-expectancy. No specific treatment is needed (Quinlivan et al., 2009, 1).

How does this condition present difficulties for the RWT? It was described on the basis of a single patient, and 'formulated' as a diagnosis on the basis of a very small number of patients. In fact, causal claims made about the disease were generally accepted after the investigation of three individuals. A brief summary of these cases will be helpful in thinking about the nature of evidence in this case.

4.5.1.1 The Discovery of McArdle Disease

McArdle, 1951, details the extensive investigation of a 30 year old man, George W. He had suffered life-long fatigue, muscle stiffness and pain on exertion. Despite his grossly normal physical examination, several subtle biochemical abnormalities were apparent. First, the level of lactate (a breakdown product of glucose) found in the blood was significantly less than expected during exercise, indicating that the patient was utilising less glucose than a normal control subject. Secondly, the patient experienced electrically silent muscle cramps[32] during exercise, indicating that his muscles were failing to contract normally.

On the basis of these results, McArdle concluded that the patient was suffering an apparently novel defect of muscle metabolism and that this defect appeared to involve a problem in the production of energy from glycogen. He therefore suggested that a defect in the glucolytic pathway was responsible. He went on to suggest that a lack of glyceraldehyde phosphate dehydrogenase (GPD) was specifically responsible, perhaps due to the action of a poison.

McArdle's publication provides a fundamentally sound clinical description of the disease. However, his causal thinking was mistaken. It is not altogether clear why this enzyme was chosen as the causative entity, as impairment of GPD results in the inability to break down glucose, rather than glycogen. Although some of the enzymes involved in glycogen synthesis and breakdown had been discovered and their properties investigated (see, for example, Cori and Green, 1943, and Green et al., 1943 for work on the properties of myophosphorylase), it is regrettably impossible to say which work in this area McArdle was familiar with. Perhaps this was the reason for his rather tentative language when presenting his findings.

The next important step in the development of a causal claim was the elucidation of the correct enzyme defect. This came about by investigation of the second known case of McArdle disease, patient D.G. He was investigated by the Mommaerts group, working in Los Angeles in the late 1950s. Two papers were published relating the features of his disease. The first was primarily biochemical (Mommaerts et al., 1959), reported an analysis of a biopsy of D.G.'s skeletal muscle. The second was largely clinical, and outlined D.G's symptoms and test results (Pearson et al., 1961). While the clinical pic-

[32] Normally, contracting muscle fibres produce an electrical signal, which can be measured. Electromyography was performed on the cramped muscles of this patient, and no electrical activity could be detected, indicating that the muscles were not contracting. Given the shortening of the muscles seen during these cramps, it appeared that the muscle fibres were 'stuck' in a contracted position due to lack of energy.

ture was extremely similar to that of George W., Mommaerts et al. disagreed with McArdle about the likely causal mechanism at work. They suggested that the characteristic symptoms were due to a lack of myophosphorylase (Mommaerts et al., 1959, 792).

This conclusion is based on the following evidence. First, in skeletal muscle samples, very little phosphorylase activity was detected (Mommaerts et al., 1959, 793-5). Second, as normal function was restored if myophoshorylase was added, it was concluded that this lack of activity was due to an absence of myophosphorylase, rather than any sort of regulatory problem (Mommaerts et al., 1959, 794). Third, the defect could similarly be bypassed if downstream products of glycogen metabolism were added (Mommaerts et al., 1959, 792-3), indicating that the metabolic problem was specific to the glycogenolysis pathway. Finally, it was noted that muscle glycogen stores were very much higher than normal (Mommaerts et al., 1959, 792), suggesting that the pathology resulted from a primary problem with glycogen catabolism, rather than one resulting from abnormal glycogen synthesis. So, in summary, the paper suggests an isolated lack of myophosphorylase was the likely causal agent for what was rapidly becoming known as McArdle disease.

The third known case of McArdle disease, patient A.D., was reported by the Schmid group, working primarily in Boston. Their work progressed in parallel with the Mommaerts group, and led to the publication of three papers in 1959 and one in 1961. Of these, two are counterparts to those published by the Mommaerts group, detailing the clinical (Schmid and Mahler, 1959) and biochemical (Schmid et al., 1959) features of patient, A.D., a 54 year old male. The third, Larner and Villar-Palasi, 1959, is a summary of measured enzyme activity from the same muscle tissue. The final paper (Schmid and Hammaker, 1961) is a short report of an hereditary study carried out on the patient's family.

Between the four papers, the group made a number of interesting observations. Chief amongst them was the discovery of post-exertional myoglobinuria, showing that muscle cells were becoming damaged by lack of energy during exercise. They also suggest a reason for the slight heterogeneity in symptoms between this patient and the other two studied by McArdle and Mommaerts. Due to the greater age of 'their' patient, they suggested a picture of slowly progressive muscle weakness

In addition to developing the clinical features of McArdle disease, the group published some observations about the biochemical lesion. They argued that the defect was due to an absence of the enzyme, rather than due to the presence of an inhibitory substance. This point stems from the observation that a 50:50 mixture of normal and diseased muscle tissue has approximately half normal phosphorylase activity (Schmid and Mahler, 1959, 2049, table V). They confirmed that liver exhibited normal glucose production from glycogen, indicating that the action of the liver form of phosphorylase was normal (Schmid and Mahler, 1959, 2054). Finally, they added further weight to the suspicion that the breakdown and synthesis of glycogen must be mediated by different enzymes.

4.5.1.2 Back to the RWT: Difference-making, not probabilities

The McArdle case therefore seems to present a problem for the RWT, in that probabilistic evidence did not seem to play a significant role in the construction of a number of (correct) causal claims, thus contradicting **RW6**. This, in turn, threatens both **RW7** and **RW12**, while also making **RW10-11** irrelevant.

This leaves us with four possible options. First, we can reject the causal claims made in this case. Second, we can reject the RWT. Third, we can treat this case as a simple exception to the RWT. Fourth, we can modify the RWT to accommodate this case.

Given that it seems inappropriately bold to reject myophosphorylase deficiency as the cause of McArdle disease on the basis of the RWT, I do not think the first option is the right one. Given Russo and Williamson's recent restatement of their thesis, such that mechanisms and probabilities are '*typically*' required (Russo and Williamson, 2010, 15), it likewise seems too strong to reject the RWT entirely. So we have two options. We could take McArdle disease as an atypical case, and one that is an exception to the RWT. This is by no means out of the question. But I think a more conceptually satisfactory alternative is to make a slight modification to the RWT to save this, and similar cases. This is to modify **RW6** and **RW7** by substituting 'evidence of difference-making' for 'probabilities'. Probabilistic evidence will, I think, be the most common form of evidence of difference-making anyway. So in most cases this modification will not affect the operation of the RWT. I develop the philosophical aspects of this modification at greater length in section 6.6. For now, I will outline the historical case for making such a modification.

I will begin by clarifying the problem. Why was probabilistic evidence absent from causal claims in McArdle disease? Put another way, what is the difference maker here? I suggest this question is best approached by considering an unusual feature of the McArdle mechanism, which instantiates clinically apparent McArdle disease in a relatively deterministic fashion. Given the absence of functional myophosphorylase, the other causal factors required to produce the clinical features of the disease - exertion, primarily - are utterly ubiquitous. Thus this single causal factor is approximately sufficient to produce the disease. Therefore, in this case, our effect (clinically apparent McArdle disease) can only arise because of the cause (absence of functional myophosphorylase). So we have clear evidence of difference-making, without any statistical investigations. This strong sufficiency of a single causal factor has parallels with the change in classificatory practices associated with the introduction of the germ theory of disease (Carter, 2003). In particular, the reclassification of clinical syndromes by their aetiologic organism, known as the aetiological standpoint, strongly resembles this move. In this way, organisms causally associated with disease became at least necessary causes of them. In the case of McArdle disease, though, the reclassification has even greater impact. Myophosphorylase deficiency becomes the sufficient, necessary and universal causal of the clinical disease.

The analagous reclassification in this case means that, by definition, we do not find myophosphorylase deficiency in the normal population. Perhaps a comparision of myophosphorylase activity in the normal versus the McArdle population could have been conducted. This would have provided probabilistic evidence in support of the causal claim. But the historical case makes no mention of such an analysis. Comparisons in the publications mentioned instead generally proceeded on a matched-control basis (see e.g. McArdle, 1951, 19-26). So instead of doing observational trials, researchers assumed that myophosphorylase deficiency was the salient difference-maker in the McArdle population. This seems a good assumption. The difference-making effect myophosphorylase deficiency exerts is strong. More-

over, the behaviour that is required for the disease to become clinically apparent (exertion) is common to the point of ubiquity. More interestingly, I suspect that as a function of this definitional sleight-of-hand, even if we were accidentally to find myophosphorylase deficiency in an asymptomatic individual, we would (probably) say they had asymptomatic McArdle disease.

There are pragmatic factors that mean that this lack of probabilistic evidence will probably persist. First, the disease is very scarce. This leads to difficulties with conducting prevalence surveys on unselected populations. Given the prevalence data on page 95, we expect to find about one case per million from an unselected population survey. There is a further difficulty with carrying out screening-type tests for the disease. Because of the wide range of mutations that can produce a functional absence of myophosphorylase, diagnosis of the disease really requires a muscle biopsy - a distinctly unpleasant procedure.[33] This means that, rather than conducting a survey of myophosphorylase activity in a range of subjects, the abnormality in clinically apparent cases is compared to an implicit norm, largely based on animal models and on occasional studies on normal volunteers. So there is no place for a survey of myophosphorylase levels.

We might also consider the role of statistical investigations within the clinical aspects of the disease. For instance, a survey of functional impairments, age of onset, and so on might be carried out. This could provide probabilistic evidence helpful in making other, related causal claims about McArdle disease. For example, we might be able to claim that "McArdle disease causes sufferers to be unable to run for 75% of buses". Once again, the extreme rarity of the disease means that these investigations will be unlikely to be carried out in a statistical manner. Instead, they usually take the form of a series of case reports, or even enumerations of all known patients. So probabilities are unlikely to play a role here either. Neither are there likely to be probabilistically-dependent causal claims about the causes of myophosphorylase deficiency. These mutations are, at present, non-modifiable - by the time of presentation at least. Once the mutation is inherited (or once a primary mutation occurs during early *in utero* development) the disease is inevitable.

This seems slightly contrived. But it has important ramifications for causal inference in medicine in general, especially when it comes to relying on case-controlled studies, or on case reports for rational causal inference. I therefore think the RWT needs to be modified in the following way. Instead of requiring probabilistic evidence, we will instead need to employ a slightly wider category of evidence of difference-making, of which probabilities will be the most commonly found sub-type.

4.5.2 Causation Without Mechanism - Hepatitis B and Liver Cancer

4.5.2.1 Introduction

Hepatitis B (HBV), an orthohepadnavirus, is a chronic viral liver infection that is causally implicated in the development of HCC.[34] While hepatitis has been known as a clinical syndrome since the prehistory of medicine, its aetiology remained unknown until the second half of the twentieth century. In this section, I

[33] This is changing slightly, as there are now PCR-based tests to identify some of the mutations. But given the scarcity of this condition, and the range of causal mutations, it seems unlikely that a simple but comprehensive test will be forthcoming.

[34] I would like to thank Vladimir Vonka for introducing me to this fascinating case

will very briefly give the history of the discovery of the causal relationship found linking HBV infection with HCC. I will suggest that, at first sight, it seems to offer an important counterexample to the RWT, as the discovery of the connection seemed to take place without a good mechanistic understanding of the manner in which the virus could cause the tumour. Does this case therefore present an example of causal discovery arising from probabilistic evidence alone?

I suggest that it does not. I will defend this claim in the following way. First, I will give a fuller outline of the story of this causal discovery. Here, I will show that while probabilistic evidence was an essential and important part of making this causal claim, it was by no means the only essential or important evidence that was actually used. I will show how mechanism, far from being absent, was an essential component of the experimental results that are conventionally construed as providing just probabilistic support. In the light of this example, I wish to make two claims about causation in medicine. The first is in support of the RWT. Examples like this demonstrate that knowledge of a causal mechanism is an essential, rather than desirable, part of understanding most causes in medicine. My second claim aims to develop the RWT by demonstrating the manner in which heterogeneous evidence can become integrated into monistic causal claims.

4.5.2.2 The discovery of HBV

While various forms of hepatitis have been known for many years, the modern conception of viral hepatitis dates from the 1930s. The most significant step arose when it was noted that a small proportion of patients (somewhere between 2 and 3%) receiving vaccination against yellow fever developed hepatitis (Findlay and MacCallum, 1937, 305). This clinical correlation was unusual. For instance, the delay of between two and six months between vaccination and hepatitis was rather longer than the known incubation period of infectious hepatitis. This raised the suspicion that some novel infectious agent was at work in these individuals. As the vaccine used in these cases was derived from human serum, and similar outbreaks were subsequently noted as a result of other instances of transfusion with human sera (zur Hausen, 2006, reviewed 244), the argument that this was a novel disease being spread between individuals by means of contaminated blood derivatives was highly plausible.

By 1947, this clinical syndrome was sufficiently well recognised to allow an aetiological classification of infectious hepatitis into hepatitis A and B (Anonymous, 1947, report on speech by FO MacCallum to the International Congress of Physicians, 1947), reflecting the recognition of underlying aetiological distinctions between the two clinical syndromes. But it was not until 1963 that the agent responsible for hepatitis B was discovered. This came about as follows.

In the early 1960's Blumberg discovered a serum antibody in the blood of a haemophiliac which reacted against an antigen found in the blood of an Australian aborigine. He named this antigen the Australia antigen (Blumberg et al., 1965, 541). Early research into this protein lead to several anecdotal reports of the infectious power of blood containing this antigen (e.g. Okochi and Murakami, 1968). Thus, a research programme developed attempting to determine the causal significance, if any, of this agent in the development of human disease. Relatively quickly, the agent was noted to be the likely cause of hepatitis B, on both statistical and mechanistic grounds (Blumberg et al., 1968). Statistically, a strong

correlation was noted between individuals suffering either acute viral hepatitis or Down syndrome[35] and the presence of the Australia antigen. Mechanistically, in addition to conventional laboratory findings, three anecdotes were exceptionally important. First, a member of laboratory staff working with the Australia antigen became Australia antigen-positive and hepatitic (Blumberg et al., 1968, 1569), suggesting that the antigen was in itself infectious. Second, another member of laboratory staff, employed immunizing rabbits with the antigen, developed antibodies directed against it (Blumberg et al., 1968, 1569-70). Thirdly, a single case was reported in which an individual, who was donating annual blood samples to the project, received a blood-transfusion which was presumably hepatitis B positive. Thus, a single case became available in which the sequential chronology of infection, seroconversion and the development of clinical disease was apparent (Blumberg et al., 1968, 1570).

However, it was not until 1970 that the origin of the Australia antigen became clear. In fact, it required the electron microscopic demonstration of VLPs in serum from hepatitis patients displaying the Australia antigen (Dane et al., 1970). It was at about this time, as a result of this demonstration and the previous work implicating the Australia antigen in viral hepatitis that the causal link between a virus and hepatitis became generally accepted.

4.5.2.3 HBV and HCC

Infectious hepatitis had been anecdotally implicated in the development of HCC since at least the 1950s (Payet et al., 1956). Early suspicions of a causal link were investigated epidemiologically, with multiple studies being conducted during the 1970s and 1980s. In general, three strategies were practised. First, cases of HCC were sought in geographical regions with known high prevalence of chronic infectious hepatitis. For example, a study in Uganda suggested a strong correlation (relative risk = 13) between hepatitis and HCC (Vogel et al., 1970). Second, populations of individuals with existing HCC were serologically investigated for patterns of antibodies or antigens indicating chronic HBV infection. For example, both Hadziyannis et al., 1970 and Sherlock et al., 1970 studied a group of Mediterranean patients who had developed HCC following hepatic cirrhosis. Many of these individuals had high serum levels of hepatitis B surface antigen. Finally, studies comparing populations with differing HBV and HCC rates were conducted. As an example, Okochi and Murakami, 1968 compared a sample in Japan (which had about a 1% prevalence of the Australia antigen in the general population) to the United States, where rates of Australia antigen were only about 0.1%. However, the precise nature of any causal relationship between HCC and HBV infection remained unclear. In fact, at least five distinct causal pathways appeared highly plausible (see diagram 4.2).

First, some authors suggested that HBV infection caused, in a direct fashion, the development of HCC (figure 4.2A). Others (e.g. Kumar and Taylor, 1973) suggested that the development of HCC predisposed individuals to develop chronic forms of HBV (figure 4.2B). Was it perhaps the case that something about having HCC caused the development of persistent antigenaemia? That is, could the

[35] Down syndrome was a marker of institutionalisation. As those living in institutions were considered at greater risk of developing infectious diseases than the population as a whole, individuals with Down syndrome had a high chance of contracting blood-borne infections such as viral hepatitis

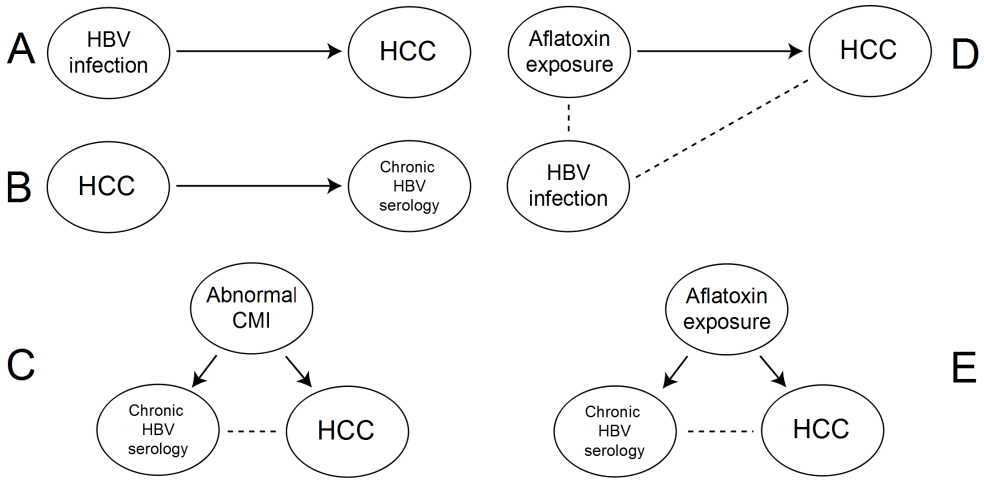

Figure 4.2: Plausible causal pathways linking HBV, HCC and aflatoxin. Arrows indicate causal relationships, dotted lines correlations. Cell mediated immunity (CMI) is the type of immune-system activity that typically acts against viral infections.

finding that individuals with HCC had high levels of antigens associated with HBV because the tumour was somehow promoting the expression of these viral proteins? Alternatively, as suggested by Dudley et al., 1972, perhaps some pre-existing condition of modified CMI could cause both the development of HCC (perhaps by the action of some unknown oncovirus) and also prevent the normal, effective response to HBV infection, with the result that markers of HBV chronicity were seen in these patients (figure 4.2C). A fourth causal hypothesis suggested was that there was no link at all between HBV and HCC. For instance, a correlation was noted between regions of high HBV prevalence and high dietary consumption of aflatoxin - a poison known to cause liver tumours (e.g. Peers and Linsell, 1973). Was aflatoxin really causing the excess incidence of HCC seen in these populations (figure 4.2D)? Finally, perhaps exposure to aflatoxin might both cause the development of HCC and also alter the normal immunological responses to HBV infection (figure 4.2E). Thus, consumption of aflatoxin might cause both the development of HCC, and also cause the characteristic serological profile of chronic hepatitis B infection (Coady, 1975).

This controversy was settled by early in the 1980s in favour of the first of the competing hypotheses. It appears that HBV infection does cause HCC.[36] How did this come about? The first really convincing epidemiological evidence for the direct causal role of HBV on HCC was a large population study of 22707 Taiwanese civil servants. Taiwan has a high incidence of HBV infection. This study showed a greatly increased risk of developing HCC given chronic HBV infection. In fact, for those infected with HBV, the annual incidence rate of HCC was 1158 per 100 000, while for those free of infection the rate was 5 cases per 100 000. This gives a relative risk of developing HCC given HBV of 233:1 (Beasley

[36] Aflatoxin also causes HCC, and appears to act synergystically with HBV. Thus individuals with both chronic HBV and high dietary consumption of aflatoxin are at very high risk of developing HCC. However, neither is a necessary factor.

et al., 1981). This is a devastatingly strong effect. Most relative risks encountered in epidemiology are at least an order of magnitude less that this. For comparison, the ten-year relative risk of lung cancer in those smoking more than 20 cigarettes per day compared to non-smokers is about 11 (Higenbottam et al., 1982). In itself, this is generally considered an impressively strong effect.

This observational claim is the first piece of vital probabilistic evidence linking the tumour with the infection. The second came about from intervention. An effective subunit vaccine for HBV, derived from human plasma, was developed in the mid-1970s (Buynak et al., 1976). It was shown to be effective in preventing HBV infection (Francis et al., 1982), and was therefore licenced for clinical use in 1981 (zur Hausen, 2006, 260). This vaccine was later supplanted by a recombinant subunit vaccine (McAleer et al., 1984). Together, these agents appears to have been effective in reducing the incidence of HCC (Zanetti et al., 2008). Thus a specific intervention acting against HBV appears to be effective in preventing its oncogenic effects. Together, these two pieces of evidence seem to offer so much support to the claim that HBV causes HCC that consideration of mechanism seems redundant. This case therefore acts as a counterexample to the RWT. That is, unless I can somehow demonstrate that mechanistic evidence plays a vital role in its formulation. In the following section, I will attempt to do just that.

4.5.2.4 HBV and HCC: Finding the mechanism

My first claim is that Beasley's epidemiological finding is compatible, more or less equally plausibly, with any of the five possible models for HCC causation given above. Neither data indicating aflatoxin exposure, nor markers of abnormal CMI, were obtained from trial subjects. Importantly, the apparent temporal relation between HBV and HCC detected is also compatible with any of the five models - assuming, that is, that some unknown time elapses between the inititiation event of any of the processes discussed and the development of detectable biomarkers. I think this is a highly plausible claim.

This observational result weakly depends upon a particular mechanistic model, implicating HBV infection in the aetiology of HCC.[37] However, the next major population finding in this area depended much more strongly on an understanding of the mechanism. This was the development of an effective HBV vaccine likely to prove effective against HCC. Two types of research were important in this. First, the development of a mechanism for HBV-mediated oncogenesis involved the discovery that the HBV genome was detectable within HCC cells (Summers et al., 1978b). This was followed by the discovery that, in some cells at least, the HBV genome was integrated with that of the host cell, leading to the expression of virally-encoded RNAs (Chakraborty et al., 1980). This opened the door to suggesting a plausible mechanism of viral oncogenesis by HBV integration and subsequent transformation of hepatocytes occurring in those with chronic hepatitis B. Most importantly, this mechanistic sequence temporally preceeded the development of macroscopic tumour formation (Shafritz et al., 1981). This process of mechanism construction was, in part, analogous to that encountered in the other cases of viral oncogenesis here described. One important difference, however, was the conspicuous role played by research on one animal model of oncogenesis. This was the woodchuck hepatitis virus, an orthohepadnavirus closely related to HBV (Summers et al., 1978a). The analogy between HBV and woodchuck

[37] For instance, in deciding to seek particular specific indicators of HBV infection in the study population

hepatitis virus was evidentially significant because this agent provided a useful animal model that could be used to understand the likely pathogenesis of HCC in a system free from the practical, ethical and legal constraints applying to experimentation on humans.[38]

In summary, before attempts at vaccination, three types of evidence mutually lent support to the causal claim that HBV was the cause of HCC. These are the clear epidemiological correlation between the virus and the tumour, dating from about 1970, the molecular mechanism implicating HBV in oncogenesis, which developed from about 1975, and investigation of the analogous case of woodchuck hepatitis virus, dating from about 1978. As one contemporary author reviewed:

> Epidemiology, molecular biology, and comparative pathology have all contributed to this inquiry. Epidemiological investigation has supplied evidence that HBV infection is associated with PHC.[39] Molecular biological studies have thrown light on the mechanism by which it may occur, while a surprising exercise in the comparative pathology of small rodents has served to reinforce the case by showing that man is not alone in this matter.
>
> (Anonymous, 1981, 1394)

So these forms of mechanistic evidence appear important in the causal story of HCC and HBV. For now I am content to say that the suggestion that this case preceded in the absence of mechanistic evidence is to simply mistake the chronology. Causally, the role played by HBV in the genesis of HCC appears to have been well accepted before the introduction of the HBV vaccine into clinical practice. More importantly, it seems highly implausible that the vaccine could have been developed without a high degree of mechanistic understanding of HBV itself. This example highlights some of the developments required of my account of mechanistic evidence, which will follow in chapter 6.

4.5.3 Probabilities and Mechanism Without Causation - Cervical Cancer and HSV

The HSV episode seems to provoke a problem with a normative interpretation of the RWT. Here, high quality evidence of both mechanisms and probabilities supported a causal claim that turned out to be faulty. In this section, I therefore outline a possible means of discovering just what, exactly, was wrong with the evidence that supported the theory that HSV caused cervical cancer. In particular, it seems that the different types of evidence in this case were not integrated together. I regard this property of integration of evidence as essential if the RWT is to be understood normatively. As such, it will form a key part of chapter 7.

Let us recap the evidence in favour of HSV as the cause of cervical cancer from chapter 3 in a manner compatible with the RWT. First, evidence of difference-making. Six pieces of difference-making evidence did the majority of the work in supporting the HSV causal claim.[40] **HSV1: venereal transmission of both diseases**. Probabilistic, observational data, such as co-variance in different patient groups, supported the claim that both HSV and cervical cancer were transmitted venereally. Sexual intercouse,

[38] I discuss the important evidential role played by analogies in greater depth in section 6.4.1.

[39] PHC stands for primary hepatocellular carcinoma, which is synonymous with HCC

[40] I will elaborate on what, exactly, characterises this evidence as evidence of difference-making in section 6.5ff.

therefore, appeared to makes a difference to the development of both HSV and cervical cancer. This was supported by **HSV2: the serological correlation** observed to exist between HSV and cervical cancer, which arose from similar, seroepidemiological data. One important additional difference-maker also arose here - **HSV3: HSV seroconversion anteceded oncogenesis**. Next, **HSV4: increased rate of HSV infections in uncircumcised men** suggested that something about circumcision prevented HSV infection. The combination of **HSV1** and **HSV4** partners with the next two difference-making claims to further suggest that it was specifically a feature of unprotected sex that made a difference to cervical cancer incidence. **HSV5: low incidence of cervical cancer in partners of circumcised men** and **HSV6: barrier contraceptives reduced cervical cancer incidence**.

Next, evidence of mechanism. Most of the difference-making evidence linking HSV, cervical cancer and sexual activity arose from similar types of investigation. These were generally large, retrospective, observational studies comparing disease rates in different populations. But the mechanistic evidence employed was methodologically and epistemically much more diverse. What exactly consitutes mechanistic evidence is something that I will explore at more length later. For now, though, it is worth noting that this evidence of a mechanism seems to fall into three groups.

First, there was a group of evidence supporting the malignant potentiality of HSV. **HSV7: HSV is mutagenic *in vitro***. As might be expected, it could therefore lead to the malignant transformation of cells in culture. Thus, **HSV8: HSV transforms some cell lines *in vitro***. So HSV causes chromosome damage, and thus can cause some sorts of cells to become malignant. These two pieces of mechanistic evidence were at the heart of the claim that HSV caused cervical cancer. Very importantly, they also depended on a vital piece of analagous evidence, **HSV9: the known, analagous role of herpesviruses in other tumours**.

Together, this group suggests that HSV could cause cancer. Did they mean that HSV could cause cervical cancer? This is addressed by two pieces of evidence specifically implicating HSV in cervical cancer. First, **HSV10: the detection of HSV in cervical cancer cells**. The specific detection of HSV in cervical cancer cells happened only once. Thus, this finding was disputed and controversial. However, it gained some support from the plausible counterclaim that existing tests for HSV were insufficiently sensitive to find the small quantities of nucleic acids likely to be found in mature cancer cells.[41] Further support was gained from **HSV11: the visual similarities between dysplastic cervical cells and cells demonstrating HSV cytopathic effects** seen in histological samples of the tumour (see 3.5.1). This issue was further complicated by the suggestion that cervical cancer could result from a model of HSV infection that did not leave residual traces of HSV in the tumour cells. This possible causal model formed an important piece of mechanistic evidence in favour of HSV, thus **HSV12: the 'hit and run' hypothesis**. Certainly, general failures of HSV detection did not seem to mechanistically exclude a causal role for HSV in cervical oncogensis.

Finally, certain events came to light that suggested the nature of the aetiological mechanism operating in study populations. These findings served to link the plausible mechanism constituted by **HSV7-12**

[41] That is, the failure of HSV detection was underdetermined

to facets of the difference-making evidence **HSV1-6**. The first of these, which served to explain the link between HSV and circumcision, was **HSV13: the isolation of HSV from smegma**. Finally, the differential susceptibility seen at various ages was explained by **HSV14: a plausible change in sensitivity of the cervix to oncoviruses in youth**.

However, the evidential story here is more complicated than it would initially appear to be. Take the first piece of difference-making evidence (**HSV1**). Not only does this suggest that sexual intercourse makes a difference to the incidence of both HSV and cervical cancer, but it also tells us some likely facts about the mechanism involved. That is, some aspect of sexual intercourse seems to play a role in both causal stories. So this claim also provides grounds for a claim about causal dependencies: HSV and cervical cancer happen or come about, in a similar way. This mechanistic claim is elaborated by other pieces of mechanistic and difference-making evidence, which further elaborate the sorts of events that are likely to be the common cause of this phenomena. So, for example, the reduction seen in both conditions with the use of barrier contraception tells us that it is something about unprotected sex, rather than sex *simpliciter*, that is causally significant. The evidence for circumcision of male sexual partners too seems to support this mechanistic claim. Therefore there are a group of, if you like, second-order, emergent pieces of causal evidence that arise when existing pieces of causal evidence are placed in context with one another. This web-of-evidence, I suggest, plays a vital role in my account of the nature by which evidence is integrated to support causal claims. I will more fully elaborate this in chapter 7.

Chapter 5

Mechanisms

5.1 Introduction

In this chapter, I aim to give an account of mechanisms that is compatible with the RWT. While mechanistic theories of causality are generally intended to be theories of causality in their own right, I here intend to do something rather different. In fact, what follows is compatible with the idea that mechanisms are simply an epistemic device that allows us to conceptualise and organise complex causal relationships. Following this line of investigation, I will give an account of mechanisms as a necessary, but insufficient, part of conceptualising causality in the health sciences.[1] To do this, I will begin by giving a short review of the current literature on mechanisms. Here, I suggest that there are a number of consensus features that arise from these disparate accounts which appear highly suited to interpreting Russo and Williamson's sense of mechanism. I will then turn to thinking about mechanisms in medicine in a more specific sense, and I will identify a number of demands that working in this area places on accounts of mechanisms. These demands will lead me to modify the consensus position in a number of ways. In particular, I will criticise the view that mechanisms can be fully characterised in terms of either causal laws or counterfactuals grounded in interventions, and I will suggest an alternative way of proceeding.

Using this modified account of mechanisms, I will then defend the use of mechanisms as tools for understanding causality in a non-reductionistic sense. Specifically, I suggest that mechanistic theories of causality, unlike regularity theories, counterfactual theories, agency theories and process theories, are capable of giving a useful analysis of causal processes that does not seek to reduce causal processes to non-causal ones.[2] Following this, I will then develop my account of mechanisms with an eye to making

[1] There might be an infinite regress here. Say we are trying to construct a mechanism between *C* and *E*, *C1 - C2...Cn - E*. However, once this is done, then the causal mechanism *C1 - C2* will, in turn, be in need of a causal mechanism too. And so on. I attempt to avoid this objection in section 5.4.3 by suggesting that, in a given context, a mechanistic model will have a fairly clearly defined level of organisation at which it bottoms-out. At this function-specific fundamental level, we will no longer be interested in the causal interactions of any more fundamental processes. Thus, while there may be a regress, it will not be an infinite one.

[2] I will term theories that seek to make this reduction *reductivist* theories of causality. For regularity theories, I include Humean regularity theories (Hume, 1975 (1748, Section VII)), Mackie's *INUS* conditions (Mackie, 1974, 60ff), as well as a number of probabilistic theories (such as Granger, 1969 and Eells, 1991). In general, these theories seem unsuited to the job of analysing causes where confounding variables or common-cause situations obtain. For counterfactual theories, I am primarily thinking of Lewis's account (Lewis, 1979; Lewis, 1986; Lewis, 2000; Psillos, 2002, 93ff). This theory suffers greatly from the difficulties that conventional accounts of world-similarity ranking face (Lewis, 1979, 47-8) when dealing with causality in medicine. For agency

it more fully compatible with the RWT. I will first suggest, and disambiguate, the role of mechanisms in giving operationally adequate interlevel explanations in section 5.4.[3] I will then suggest, as a consequence of this interlevel behaviour that mechanisms need to be carefully specified in terms of their function if they are to play a useful role in forming RWT-type causal claims. I will do this in section 5.5. That is, we need to carefully consider exactly *what* a particular mechanism is *for* when we attempt to construct it. I will then suggest, in section 5.6, that most medical mechanisms, if correctly functionally specified, will display a characteristic branched structure, and that an important part of thinking about causes in medicine is to decide just which causal 'track' is taken in a particular instantiation. This will form the basis for the final part of this chapter, which is to describe just how this kind of functionally specified, branched mechanism can accommodate causation by prevention or by omission, both of which are generally difficult using other theories of causality.

This account of mechanisms then requires an account of mechanistic evidence. This will follow in the first half of chapter 6. I will then turn to thinking about difference-making in the second half of that chapter. I will return to mechanisms, and their relationship to difference-making, in chapter 7, where I will give an account of the expanded RWT.

5.2 Mechanistic theories of causality

We might begin with Williamson's recent definition of mechanistic theories of causality in general:

> A mechanistic theory of causality holds that... two events are causally connected if and only if they are connected by an underlying physical mechanism of the appropriate sort.
>
> (Williamson, 2010, 2-3)

I suggest this applies to the original formulations of all the theories of causality that follow. As I previously mentioned, though, I think that mechanisms do not tell us the whole story about causality in medicine. I suggest, in line with the RWT, that describing causal relationships in terms of mechanism and the relevant evidence used to construct them is a necessary, but insufficient component of assessing causation. For the time being, then, I will take the conceptual components of these accounts of causality as my focus, and I will say very little in specific terms about how they fare as theories of causality in their own right. In any case, I think there are two simple objections to treating all causes as simply mechanisms. The first is the difficulty of dealing with causation by omission or by absences, while the second relates to the difficulty of accounting for the types of evidence that are actually employed in scientific practice to form and assess causal claims. I think that these problems are closely related, and I suggest just an account that combines mechanistic with other forms of evidence can be formulated - and used to address the problems of negative sorts of causation - later. With that proviso in mind, I will give here

theories, I'm concerned with Menzies's and Price's (Menzies and Price, 1993, 187) theory, which seems unable to account for situations where natural experiments seem to provide us with useful causal knowledge. I deal with process theories of causality in section 5.4.1.

[3] To distinguish it from reductivism, I shall call theories that define causality in terms of microphysical interactions *reductionistic* theories of causality

a short introduction to five recent mechanistic theories of causality, which have distinct similarities. A little terminological disambiguation is in order at this stage. First, I intend to refer only to that recent group of mechanistic theories that are referred to as the New Mechanistic Philosophy (NMP). I do not intend to address mechanisms in the reductive, Cartesian, historical, sense (see e.g. des Chene, 2005). Second, when addressing these recent theories, I will restrict my discussion to those arising from the *complex-system* tradition. I will neglect, for the time being, a second, reductionist, group of mechanistic theories of causality. These are the process theories, reviews of which can be found in Psillos, 2002 and Williamson, 2010, 4ff. Later, in section 5.4, I explain why I regard these theories - Salmon's mark-transmission (MT) theory (Salmon, 1984) and Salmon and Dowe's CQ theories (Salmon, 1998; Dowe, 2000) - as unsuitable for the analysis of causality in medicine, and likewise suggest that the sort of epistemic tools that these theories licence are inappropriate for the task of representing causal relationships in medicine. With those distinctions under our belts, I will outline some NMP-type, complex-system theories of causality.

5.2.1 Machamer, Darden, Craver

The first of these accounts is that of Machamer, Darden, Craver (MDC). Their account is explicitly based upon studies in the biological sciences, most notably neuroscience, and has been extensively developed over many publications on the subject (see, for example, Machamer et al., 2000; Craver, 2001; Craver, 2002; Darden, 2002; Machamer, 2004; Craver and Darden, 2005; Darden, 2005; Darden, 2006 and Craver, 2007). However, their account of mechanisms begins with an highly succinct definition:

> Mechanisms are entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions.
>
> (Machamer et al., 2000, 3)

Let us analyse these claims in detail. First, MDC claim that mechanisms are - ontologically - constituted by both **entities** and **activities**. Entities are the things, the physical matter, which participate in the causal process via their properties. Examples might include viruses, cells, proteins, nucleic acids, organs and so on. Activities, on the other hand, are the 'producers of change' (Machamer et al., 2000, 3). Entities thus act on other entities via their activities. So a protein might *bind* a ligand. Or a virus might *infect* a cell. These causal verbs are the activities, and their actions between a series of entities is responsible for producing the function of the mechanism. Description of both entities and activities are therefore required to give an ontically adequate account of a mechanism. Thus they reject accounts of causality that attempt to reduce causality to an exclusive ontological property of either substances or processes (Machamer et al., 2000, 4-5). Craver later proposed that this dualistic conception of entities and activities brings epistemic and methodological advantages when compared to an account that prioritises either alone:

> Calling them "activities" rather than "capacities"... emphasizes the way that mechanisms work as opposed to the way that a set of parts has the capacity to work or is disposed to work. This shift of gestalt is recommended and defended...

(Craver, 2001, 58, footnote 2)

Thus MDC's account rejects both process and substantivalist ontologies, instead placing entities and activities on an separate-but-equal ontological footing. I agree that these non-ontological advantages are important, and I employ them as part of my own account of mechanism. However, MDC do not offer a stronger defence of this dualism than similar abductive arguments from epistemology. For instance, they suggest that the apparent independence of activities and entities as subjects of enquiry suggests this ontic dualism. In defence of this stance, Machamer later wrote:

> From the fact that activities have measures that may be applied to them, and from the fact that they may be identified independently of any particular entities that are acting, we might infer that they have some independent status as referents. This is presumably what we need in order to attribute ontological status. Nevertheless, such general characterizations do not begin to come close to a definition.

(Machamer, 2004, 29)

Now this ontological argument, I think, has convincing epistemic consequences. It is the case that we quite often identify causal activities without identifying entities to which those activities relate. For example, Lucké (see section 1.1) was able to very clearly describe the clinical features which resulted from the activities of an unknown virus, without really understanding how these activities related in detail to any particular infectious agent. Likewise, we may find and describe an entity without necessarily knowing its activities. For instance, we might discover a viral genome without knowing how, or even if, it acts to produce disease. I develop this epistemic and methodological argument for dualism further in section 5.3.

The next feature of MDC's account deals with the manner in which activities and entities are arranged. MDC claim that mechanisms are **organised** in some way. As they suggest, mechanisms are not simply aggregates of things. Instead, the activities and entities constituting a mechanism are distinctively organised in some way. Mechanisms themselves are characterised by the 'spatial and temporal organization of the[ir] component parts' (Craver, 2007, 135–6).[4] By way of effecting this differentiation of mechanisms from aggregates, MDC identify a special kind of organisational property that is characteristic of mechanisms: **active organisation**. Active organisation is a description of the manner in which the components of a mechanism interact together. As Craver puts it, active organisation is characterised by the arrangement of 'activities of and interactions among the entities in the working mechanism' (Craver, 2007, 136).

This property arises from the particular spatial and temporal patterns of organisation in the mechanism in question (Craver, 2001, 60ff). That is, the active organisation of a mechanism comes about from the particular organisation of its components in terms of, say, their size, orientation, shape, duration, manner of contact and so on (Craver, 2007, 137–8).

[4] Craver, 2007 does a great deal to disambiguate this distinction between mechanisms and aggregates along lines originally suggested by Wimsatt (1997).

For mechanisms, the term organisation thus refers to two different, interrelated, properties. The first is the particular pattern of organisation that the components of a specific mechanism displays. The second - active organisation - arises from this particular pattern and is the source from which claims about causal function can be made for that mechanism. In order to give an adequate description of a mechanism, we therefore need to consider two kinds of organisation. We first need to give a description of how the components of that mechanism are organised in terms of shape, spatial arrangement, temporal sequence, manner of interactions and so on (Craver, 2007, 136–7). For example, the manner in which a drug interacts with a receptor may be a question of the morphologies of both the interacting entities, or the organisation of particular enzyme complexes, or of the spatial arrangement of parts of the interacting entities. However, in addition to describing how the components of a mechanism are organised, we also need to give a description of how these organisational properties give rise, together, to the property of active organisation.

Thus, for MDC, active organisation is a property that is distinctively mechanistic. While other kinds of organisation, such as spatial arrangement of parts, can happily be a characteristic feature of non-mechanistic aggregates, mechanisms alone have the property of active organisation. Without active organisation, we would no longer have a mechanism. As a consequence, performing interventions that interfere with active organisation - by rearranging, substituting, disaggregating or reaggregating parts - typically breaks, or otherwise alters, the function of a mechanism. In contrast, the parts of an aggregate can be rearranged without such functional disruption.

Active organisation, as a general property of mechanisms, gives rise to the feature of **regularity**. MDC suggest that regularity deals with '... the typical way the mechanism runs from beginning to end' (Machamer et al., 2000, 3). That is, a degree of regularity between a cause and an effect arises in a given mechanism because of the existence of active organisation in that mechanism. When a mechanism is suitably organised there therefore exists a **productive continuity**, between each of the particular links of the mechanism, linking cause and effect. In order to provide an adequate description of a mechanism, we need to demonstrate the existence of this productive continuity as an unbroken chain of causal relations linking cause and effect:

> Complete descriptions of mechanisms exhibit productive continuity without gaps from the set up to termination conditions.
>
> (Machamer et al., 2000, 3)

If this organisation is disrupted then the connection will be lost. Thus it is something about the sequence of causes at play in a mechanism that require us to make the distinction between aggregates and mechanisms. In other words, mechanisms can only exhibit their behaviour when organised. If this organisation is lacking then the instantiation of a cause in the mechanism will (generally) give rise to different, or no, effects.

Therefore the overall relationship between a cause and an effect that a mechanism describes and the linked sequence of causal relationships that constitute it means that this account of causality is **non-reductivist**. In fact, MDC adopt a distinctly agnostic attitude regarding any more reductivist account of

causality. Theirs is an account of causality that neither seeks to reduce causal to non-causal concepts (as further outlined in Craver, 2007, 64), nor to seek the secret connexion that Hume could not find.[5] This is not always the case with NMP-type theories of mechanism, as I will address later. What this theory does purport to do, though, is to allow the analysis of complex causes in convenient terms. For example, it might permit an investigator to understand the manner in which a cause operates such that they can give an explanation. Or perhaps this process of conceptualising causes might be performed for reasons of exploiting the kinds of special features that we expect of causal relationships.[6]

As a related issue, MDC's understanding of causes does not simply rely on relating the operations of causes to causes operating at the most fundamental level. Their account is therefore **non-reductionistic**. Instead of thinking about causes as encountered in scientific practice as being explainable only in terms of their relation to, say, micro-physical causal interactions, MDC instead value higher-level phenomena as genuinely explanatory. In other words, for MDC, mechanisms are **interlevel** (this position is most strongly developed in Craver, 2005).

However, moving to using this theory of causality brings some difficulties that relate to what, exactly, we mean by a mechanism. There is a vital clarification that needs to be made here between mechanisms, models of mechanisms and representations of these models. I take the definition given above to speak to claims about the mechanism as it exists in the world. That is, I think these claims about entities, activities, organisation and so on are about the ontology of mechanisms themselves. However, when it comes to thinking about understanding and exploiting causes we move to discussing features of our mental models of mechanisms. That is, claims about the performance of MDC's account in an operational sense will involve thinking about features of these causal models, rather than the causes as they exist in the world. I think that models of mechanisms have important differences from ontological mechanisms. I will address some of these below, and I will also return to this issue when I come to discuss Glennan's account of mechanisms in section 5.2.2.

But for now, I can identify aspects of the model/mechanism distinction in MDC's work. For example, they suggest that our interpretation of a mechanism (i.e. as a mechanistic model) is contextualised:

> Functions are the roles played by entities and activities in a mechanism. To see an activity as a function is to see it as a component in some mechanism, that is, to see it in a context that is taken to be important, vital, or otherwise significant.
>
> (Machamer et al., 2000, 6)

I take this sort of argument to imply that if we wish to suggest a mechanism as a potential explanation of a phenomena then our descriptions of that mechanism must take account of the context within which it operates. In particular, we will have to specify just which one of the many possible behaviours of that mechanism we are seeking to employ in our explanation. For example, the heart both pumps

[5] 'It appears that, in single instances of the operation of bodies, we never can, by our utmost scrutiny, discover any thing but one event following another, without being able to comprehend any force or power by which the cause operates, or any connexion between it and its supposed effect.' (Hume, 1975 (1748, 12.58)

[6] c.f. Russo and Williamson's claims about cognitive and action-oriented goals, which are reviewed in section 4.2.

blood and produces heart sounds. If we are trying to give a mechanistic explanation for circulation, we will seek to describe the pumping behaviour, rather than the sound-producing behaviour, of the heart as part of our mechanistic explanation. MDC describe this kind of consideration as **functional specification**. I return to *why* there is a need for this sort of practice in section 5.2.6. However, the *how* is worth describing here. In practice, this functional specification is performed by describing, in some detail, the set-up and termination conditions for the mechanistic model. There are a number of provisos that MDC have made about these which need not concern us too much at this stage (e.g. that these conditions will be relative to both a set of idealised background conditions and to the initial states of the entities and activities. Likewise a number of implicit *ceteris paribus* conditions may well be crucial to seeing how the mechanism will actually go (Craver, 2001, 11ff)). However, the general principle is quite straightforward.

Both set-up and termination conditions are defined at explanatorily privileged endpoints (Craver, 2001, 11-2). For instance, if we wish to explain why an individual developed a disease, we might select a set of set-up conditions that represent a normal, physiological state, while our termination conditions would represent some kind of description of the disease of interest. On the other hand, if we were to seek a causal explanation of the success of a treatment for this same disease, our set-up conditions would be descriptors of the disease, and our termination conditions descriptors of the cured or palliated state that (hopefully) obtains following treatment. Thus set-up and termination conditions depend upon the functional specification of the mechanism in question. Thus the causal structure represented by a mechanism is highly context-dependent. We can only specify these conditions once we give an account of what the mechanism in question stands for. They are therefore a feature of mechanistic models, rather than mechanisms:

> To give a description of a mechanism for a phenomena is to explain that phenomenon, i.e. to explain how it was produced.
>
> (Machamer et al., 2000, 3)

This context-dependent aspect of mechanistic models means that mechanistic models of a particular mechanism are liable to change with circumstances. This means models of mechanism are dynamic; they can change over time.

> ...one should not think of mechanisms as exclusively mechanical (push-pull) systems. What counts as a mechanism in science has developed over time and presumably will continue to do so.
>
> (Machamer et al., 2000, 2)

This, as I interpret it, supports MDC's claim that their mechanistic metaphysics arise from scientific practice:

> In many fields of science what is taken to be a satisfactory explanation requires providing a description of a mechanism. So it is not surprising that much of the practice of science can be understood in terms of the discovery and description of mechanisms.

(Machamer et al., 2000, 1-2)

5.2.2 Glennan I

Glennan has produced two differing positions regarding mechanisms. His first, which might be described as a *mechanical* theory (Glennan, 1996) relies on interpreting mechanisms in terms of causal laws, whereas the second, *mechanistic* theory (Glennan, 2002) explicates causal interactions between parts of a mechanism in terms of Woodward's account of 'invariant, change relating generalizations' (Woodward, 2000). I will therefore introduce his earlier formulation here, and I will interleve an account of Woodward's position on mechanisms before I discuss the second.

Glennan begins by arguing, contra Hume, that there is more to causality than regular succession. He suggests that a mechanical theory of causality instead explains the necessity that seems to distinguish laws from accidental generalisations in an unproblematic manner:

> When I claim that some event causes another event, say that my turning the key causes my car to start, I do not believe this simply because I have routinely observed that turning the key is followed by the engine starting. I believe this because I believe that there is a mechanism that connects key-turning to engine-starting. I believe that the key closes a switch which causes the battery to turn the starter motor and so forth. Furthermore, this is not a "secret connexion". I can look under the hood and see how the mechanism works.

(Glennan, 1996, 50)

That is, a causal relationship should be characterised by both behavioural regularity and the existence of an observable mechanism capable of linking cause with effect. So a causal explanation of a correlation between C and E will require the demonstration of a mechanism, such as C–$D1$–…–Dn–E, connecting C and E. He goes on to offer a suitable definition of mechanism:

> **(M)** A mechanism underlying a behavior is a complex system which produces that behavior by of the interaction of a number of parts according to direct causal laws.

(Glennan, 1996, 52)

That is, the individual causal linking in the mechanism C–$D1$–…–Dn–E should be understood as instances of direct causal laws which govern the interactions of the constituent parts. The overall behavioural correlation seen between C and E is then explicable in terms of the bundle of individual causal laws that govern the interactions between the components of the mechanism:

> …a mechanism is a collection of parts, in which the behavior of the aggregate stems from a series of local interactions between parts.

(Glennan, 1996, 56)

Glennan further suggests that, characteristically, mechanisms are behaviourly *polymorphous* (Glennan, 1996, 52). For example, combustion engines both move drive-shafts and produce heat. Glennan

claims that we can give legitimate mechanistic explanations of either of these behaviours by appropriately decomposing the mechanism into its relevant parts, depending upon what is to be explained. This means that, depending upon what behaviour it is that we wish to explain, we can analyse a given mechanism in a number of different ways to produce a range of mechanistic models. However, such modelling practices are negatively constrained by ontic features of the mechanism itself: in order for a mechanistic model to provide a good description of a behaviour its components must be 'objects' (Glennan, 1996, 53) that are 'really there' (Glennan, 1996, 52). So this context-dependent decomposition is not intended to undermine claims of realism about the parts of the mechanism. Now these two facets of Glennan's account - mechanism-for-behaviour and behavioural polymorphism - are very useful for understanding the distinction between mechanisms and mechanistic models, and I therefore take them up in section 5.2.6.

Glennan suggests that the manner of interactions between parts of a mechanism are governed by causal laws, and that these causal laws should be understood as generalisations or universal propositions which support counterfactuals (Glennan, 1996, 54). While this is clearly non-reductivist, it is rather different from the non-reductivist approach to causality suggested by MDC. While MDC were non-reductivist in the sense that they did not specify a particular prior notion of causality at all, Glennan instead uses one type of causality (that of direct causal laws) to support his mechanical view. However, like MDC, this makes his account to a degree circular: we can explain the lawlike characters of mechanically explicable laws by reference to the mechanisms that underlie the laws themselves (Glennan, 1996, 63). Again, like MDC, this circularity regarding causality is not vicious; even though we can only really explicate causes in terms of other causes we can still provide useful explanations of complex causal situations by reference to simple causal components that are easy to understand and control. It might be argued that this position tends towards regress, in that we may well have a problem explaining the lawlikeness of fundamental causal laws. This means that his theory offers no attempt to explain causality in fundamental physics (Glennan, 1996, 50). However, he suggests that the best solution to this dilemma is a dichotomous understanding of causation in fundamental physics vs understanding causation everywhere else. In other words, the truth-makers of the majority of causal claims (which happen at non-fundamental levels) under Glennan's theory are unproblematic and intuitively appealing, but incoherent at a fundamental level. This dichotomous understanding is permissible, according to Glennan, because we need not worry too much about reducing all causal claims to the operations of physical systems. He supports this argument with two pieces of empirical evidence from scientific practice. The first is that the vast preponderance of causal talk solely concerns higher-level causal claims:

> Causal statements are typically statements about events regulated by mechanisms, and mechanisms are complex, higher level entities. Only when we talk about interactions governed by fundamental laws does causal talk become problematic.
>
> (Glennan, 1996, 68)

The second, based upon a reading of quantum entanglement, suggests that there may very well be no causality governing fundamental interactions anyway (Glennan, 1996, 66-7). So the results of these

two pieces of evidence is to suggest that causal claims about fundamental phenomena are generally irrelevant to scientific practice, and that discussions of their nature may, in any case, be incoherent. Given this proviso about their scope, how does Glennan suggest that these apparently non-fundamental causal laws are constituted? First, Glennan requires that part-part interactions of a mechanism must be governed by 'direct causal laws' (Glennan, 1996, 55). By this, he means to suggest that the causal laws themselves must give an account of just how two phenomena are causally connected. This excludes lawful generalisations that explain two phenomena as the actions of a common cause. As he suggests:

> ...it is a lawful generalization that night follows day, but certainly day does not cause night. Rather, the onset of day and of night are events which are both caused by the earth's rotation. Relations between parts must be governed by causal laws because otherwise the parts could not be said truly to interact.
>
> (Glennan, 1996, 55)

5.2.3 Woodward

Woodward first set out his position on causality in Woodward, 2000 and Woodward, 2002, and later substantially developed it in Woodward, 2003. This mechanistic theory of causality essentially aims to overcome a number of difficulties that arise from relating causal explanations to the operation of laws. His essential idea is this: that causal explanations are (and should) be done by relating phenomena to the operations of invariant generalisations, rather than laws. This has a number of technical advantages. For example, as Woodward suggests, the exact degree of invariance - across a particular range of background conditions, for instance - can meaningfully vary. This grants invariant relationships a rather different status to those relationships that are governed by laws, which are generally suggested to take the form of exceptionless regularities of universal scope. These law-governed generalisations are generally assumed to be capable of being simply dichotomously true or false. In contrast, using invariance gives us the ability to speak to the truth-value of relationships in a more granular manner. For example, we can suggest that a particular causal relationship is likely to obtain over a particular range of conditions. This localisation is something that is not, in general, associated with truth-makers of relationships governed by laws. So what is this property of invariance? Woodward claims that generalisations are *invariant* if they are both change relating and stable. *Change relating* means that these generalisations speak of the relationship between two event-types, such that the instantiation of one event-type brings about some sort of difference in the second event-type. *Stability* here also has a specific meaning. As Woodward suggests, stable generalisations are those that are robust under *interventions*. This relation of causes to interventions is the main characteristic of Woodward's account. He begins by defining an intervention:

> An intervention is an exogenous causal process that changes some variable of interest X in such a way that any change in some second variable Y occurs entirely as the result of the change in X.
>
> (Woodward, 2000, 199)

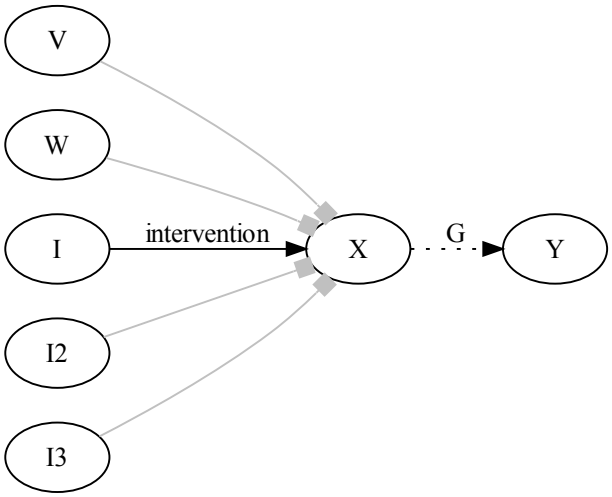

Figure 5.1: Woodward's intervention condition M1. Here, *X* is changed by intervention *I*. *G* stands for the causal relationship, if any, between *X* and *Y*. M1 excludes any other causes (*V*, *W*, *I2* and *I3*) from changing *X*.

He imposes certain conditions on these interventions. Take an intervention *I* on the variable *X* (Woodward, 2000, 201). First, *I* should change *X* in some way, and this change should be entirely due to *I* (M1). Second, according to postulated relationship between *X* and its effect, *Y*, the change brought about by *I* on *X* should change *Y*. That is, the value of *Y* under intervention *I* on *X* should differ from the value of *Y* when no such *I* had taken place (M2). Third, *I* should change *Y* only via *X*. That is, there should be no effect of *I* that has an effect on *Y* via any route a) that is not an effect of *X*, and b) that are not intermediate between *I* and *X* that have no effect on *Y* in the absence of *X* - that is, effects of *I* of this type may not be sufficient causes of *Y* (M3). Finally, (M4) suggests that *I* should not correlate with any other cause of *Y* except *X*, and those cases mentioned in (M3).

I find these adequacy conditions rather confusing. I have therefore attempted to represent them diagrammatically below in figures 5.1-5.4.[7]

Thus, for Woodward, any causal process that meets conditions M1-4 counts as an intervention. This means that, for example, natural experiments count as interventions (Woodward, 2000, 201).[8] Thus

[7]Here, I am trying to take one of the aspects of the normativity so frequently encountered in the mechanisms literature seriously (c.f., for instance, Bechtel and Abrahamsen, 2005, 426ff). I think that graphical representations of mechanisms provide a substantial benefit, over textual representations of the same phenomena, when it comes to understanding the operation of complex systems. With that intuitive claim in mind, I have lately been working on a range of tools for building mechanistic representations in a way that is operationally convenient. That aside, Woodward gives some adequacy criteria for mechanistic representations, which I will mention at the end of this section.

[8]'[this] characterization makes no essential reference to human activities or to what human beings can or can't do. A causal process that does not involve human beings at any point will qualify as an intervention as long as it meets conditions M1-4. Indeed, it is precisely this sort of possibility one has in mind when one talks about a "natural experiment" (Woodward, 2000, 201)

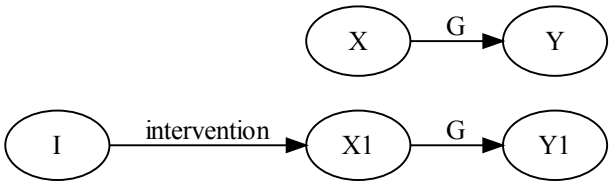

Figure 5.2: Woodward's intervention condition M2. This assumes that the relationship condition, *G*, is something like $X \hookrightarrow Y$ & $X1 \hookrightarrow Y1$. Thus, *X* usually causes *Y*. But intervention *I* changes *X* to *X1*, thus instantiating *Y1* instead.

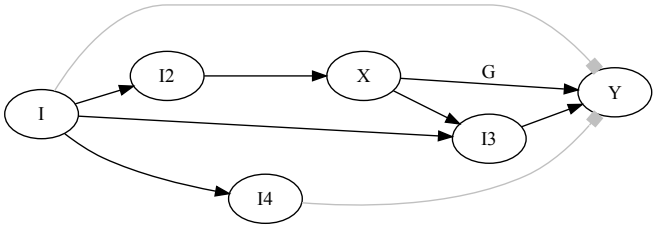

Figure 5.3: Woodward's intervention condition M3. *I3* is allowable because it is an effect of *X* (M3a). *I2* is likewise allowable, because it is an intermediate between *I* and *X* (M3b). However, neither the direct effect of *I* on *Y*, nor *I4* are allowable, because they are neither intermediate between *I* and *X*, nor intermediate between *X* and *Y*

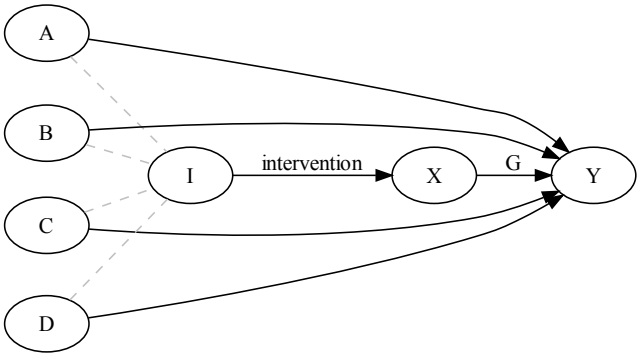

Figure 5.4: Woodward's intervention condition M4. There should be no causes of *Y* (in this example, *A*, *B*, *C* and *D*) that correlate with *I* except *X*.

it is the sort of causal structure, and not the question of agency, that makes a process an intervention. Interventions on X, and their effects on Y, can therefore tell us that there is a causal relationship between X and Y. His account is, in common with the other theories of mechanisms, non-reductive, as causal talk features in the definition of interventions, which in turn are used in Woodward's definition of causation.

Anyway, interventions provide grounds for causal relationships in the following way. Given a particular causal system, a range of interventions can be performed on it. This range of interventions allows the establishment of a number of counterfactual dependencies that tell us just how the causal system behaves:

> ...exhibition of patterns of counterfactual dependence of a special sort, involving active counterfactuals - counterfactuals the antecedents of which are made true by interventions. Only invariant generalizations will support active counterfactuals...
>
> (Woodward, 2000, 199)

This has the important result that Woodward's theory is therefore one that speaks only to the generic case. We cannot, quite clearly, perform a range of interventions on a causal system and determine a pattern of counterfactual dependence if that system only comes about once. However, assuming that we have a system which is a suitable one for this sort of repetitious approach, Woodward tells us how we can create mechanistic representations that allow us to put all these elements together to form something that can operationally be employed to understand complex causes:

> **(MECH)** a necessary condition for a representation to be an acceptable model of a mechanism is that the representation (i) describe an organized or structured set of parts or components, where (ii) the behaviour of each component is described by a generalization that is invariant under intervention, and where (iii) the generalizations governing each component are also independently changeable, and where (iv) the representation allows us to see how, in virtue of (i), (ii) and (iii), the overall output of the mechanism will vary under manipulation of the input to each component and changes in the components themselves.
>
> (Woodward, 2002, S375)

Now I have some historically motivated concerns about this definition. Taking point *iii)* in particular, it is very often not the case that a claim for interventional independence of each component of a causal mechanism is even plausible. For example, a reading of the viral oncogenesis cases from chapters 2 and 3 suggests that many variables will not be amenable to such modification. In specific terms, what happens when our participating causal variable is something like a virus species? What does it mean to change this under intervention? In general terms, it may well be that some variables, such as X, cannot take different values. It cannot therefore be changed by I such that it alters Y. Now this sort of concern does not, for me, licence totally discarding Woodward's theory. What it does suggest, though, is that this sort of interventionist approach does not tell us the whole story about causality. I therefore will take up several of Woodward's suggestions about how we can determine the structure of complex causes later.

5.2.4 Glennan II

Glennan's second mechanistic scheme really differs in just one respect from his first account. Instead of explicating the causal connections between mechanistic components in terms of causal laws, this revised account employs Woodward's notion of change-relating invariant generalisations (CRIGs):

> **(M)** A mechanism for a behaviour is a complex system that produces that behaviour by the interaction of a number of parts, where the interactions between parts can be characterized by direct, invariant, change-relating generalization.
>
> (Glennan, 2002, S344)

Again, Glennan carefully characterises the relationship between behaviours and mechanisms. First, mechanisms have an external behaviour (Glennan, 2002, S347) which can be described. This external behaviour is underwritten by a mechanism that is capable of accounting for its features. Thus mechanistic models need to describe both the external behaviour of a mechanism, and the manner in which the underlying mechanism in the world is supposed to account for that behaviour:

> **(MM)** A mechanical model is a description of a mechanism, including (i) a description of the mechanisms behavior; and (ii) a description of the mechanism which accounts for that behavior.
>
> (Glennan, 2002, S347; see also Glennan, 2005, 446 for a slightly revised version)

This suggests that there are two differing aspects of providing an adequate mechanistic model that we might seek when looking for historical support for this account of mechanisms. First comes the description of the behaviour of the mechanism. Glennan outlines just what this involves:

> ...the concept of a mechanism's behavior generally presupposes a concept of normal functioning. When one describes the behavior of a mechanism, one describes how it will behave if it is not broken.
>
> (Glennan, 2005, 448)

I think this is a critical property of adequate mechanistic description of causal relationships in medicine. In general, a pathological process is defined by a contrastive relation with the relevant physiology. In other words, our concept of a pathological mechanism for a particular set of symptoms or behaviours presupposes a difference between that mechanism and what would (counterfactually) happen in a healthy state. I will return to this when I discuss functional specification of medical mechanisms (in section 5.5) and again when I introduce the concept of branching (section 5.6). For now, though, I will simply assert that, because of the behavioural polymorphism of mechanisms, we will need to specify carefully just which behaviours we are attempting to explain by our mechanistic models. This is critical for a second reason, which Glennan also raises:

> ...there is a one-many relationship between behavioural and mechanical descriptions. This is because the same behavior can be produced by different mechanisms.

(Glennan, 2005, 449)

So not only is it the case that a given mechanism will tend to have more than one (mechanistically explicable) behaviour, but also that a given behaviour may well have associated with it a number of possible causal mechanisms. Dealing with this unclear correspondence between behaviours and mechanisms is a key task for my account of causality, so I will return to this issue in chapter 7.

5.2.5 Bechtel

Bechtel[9] (Bechtel and Richardson, 1993) set out a theory of mechanisms that is based on a study of the life sciences.[10] In brief, Bechtel uses an interlevel concept of mechanisms in order to give meaningful non-nomic explanations. As such, his account shares much with those that came later. However, it is distinctive in respect of the type of things that participate in a mechanism. Rather than activities and entities, interactions or CRIGs, he characterises mechanisms as composed of component parts that operate on one another:

> A mechanism is a structure performing a function in virtue of its component parts, component operations, and their organization. The orchestrated functioning of the mechanism is responsible for one or more phenomena.

(Bechtel and Abrahamsen, 2005, 423)

Bechtel uses this particular terminology because he wishes to prioritise the conceptual primacy of the parts that participate in the mechanism. Physical parts, Bechtel claims, are the essential constituents of mechanisms, and the operations of these parts depend fully upon them:

> We use the term *operation* rather than activity because we want to draw attention to the involvement of parts; for example, enzymes operate on substrates so as to catalyze changes in the substrates.

(Bechtel and Abrahamsen, 2005, 423)

So unlike MDC, Bechtel wishes to place mechanisms on a more substantivalist footing. Thus the causal interactions that together constitute the mechanism are properties of the physical parts of that mechanism. Groups of these inter-part interactions constitute the component operations that can then be used to describe the mechanism:

> Each component operation involves at least one component part. Typically there is an active part that initiates or maintains the operation (and may be changed by it) and at least one passive part that is changed by the operation. The change may be to the location or other propert(ies) of a part, or it may transform it into another kind of part.

[9] In the interests of terminological simplicity, I will refer to this position as that of Bechtel, and I will neglect the specific contributions of his collaborators.

[10] 'In this paper we are targeting explanations in the life sciences. We are not claiming all scientific explanations involve appeals to mechanisms or even that in biology all explanations take the form of identifying the responsible mechanism. The claim is far more modest – that in many cases in the life sciences, and potentially in other sciences, the quest for explanation is the quest for a specification of the appropriate mechanism.' (Bechtel and Abrahamsen, 2005, 422)

(Bechtel and Abrahamsen, 2005, 424)

And it is the orchestrated and organised character of these component operations that makes mechanisms work. As with the other NMP accounts, this organisation is characteristically complex.[11] Likewise, Bechtel aims to differentiate the mechanism in the world from the kinds of epistemic activities that can be performed with knowledge of that mechanism. In particular, mechanistic explanations can easily be incorrect, or even absent, without challenging the actual causal operations of the ontological mechanism:[12]

> ...mechanisms are real systems in nature...but it is crucial to note that offering an explanation is still an epistemic activity and that the mechanism in nature does not directly perform the explanatory work...This is particularly obvious when one considers incorrect mechanistic explanations - in such a case one has still appealed to a mechanism, but not one operative in nature. Another way to appreciate the point is to note that in many instances the mechanism in question was operative long before scientists discovered the mechanism and invoked it to explain the phenomenon. Thus, since explanation is itself an epistemic activity, what figures in it are not the mechanisms in the world, but representations of them. Bechtel has also provided a number of detailed accounts of how these epistemic models are constructed from ontological mechanisms account of the kinds of decomposition and complexity (Bechtel, 2001, 2008).

(Bechtel and Abrahamsen, 2005, 424-5)

Now these mechanistic models are therefore representations of the operation of a mechanism in the world, and must therefore describe a number of features of it if they are to be explanatorily adequate. These include the parts and operations of these parts, the manner in which these parts and operations work together, and the overall organisation of the mechanistic system that is responsible for producing the phenomena of interest:

> A model of a mechanism describes or portrays what are taken to be its relevant component parts and operations, the organization of the parts and operations into a system, and the means by which operations are orchestrated so as to produce the phenomenon.

(Bechtel and Abrahamsen, 2005, 425)

They also make an interesting suggestion about levels. While it is the case that Bechtel's mechanisms are interlevel, it is generally the case that the interactions in a mechanism are usually between parts that are of a similar magnitude. This means that Bechtel seems partially to define the concept of levels by reference to the typical operation of mechanisms:

[11] 'Operations can be organized simply by temporal sequence, but biological mechanisms tend to exhibit more complex forms of organization...mechanisms may involve multiple levels of organization.' (Bechtel and Abrahamsen, 2005, 424)

[12] I interpret this as a claim about mechanisms versus mechanistic reasoning. This means that criticisms of their position that rely on demonstrating faulty mechanistic reasoning do not undermine their assertion that it is mechanisms that are responsible for effects. I defend a similar position in what follows.

> Models posit parts of a particular sort within a system and operations performed by these parts. The task in developing a model is to identify these parts and ascertain what they do and how they interact to produce the phenomenon of interest. Typically, a part found within a system will interact with other parts of roughly the same magnitude, and this cluster of interacting parts will constitute a level.
>
> (Bechtel, 1994, 15)

5.2.6 Consensus features of the NMP

While there are a number of vital differences between these accounts, I would like to extract some consensus features of mechanisms.[13] First, all five accounts claim that mechanisms are composed of real **entities that causally interact**. Whether this is described in terms of entities and activities (MDC), components and laws (Glennan I), components and CRIGs (Woodward; Glennan II) or components and functions (Bechtel), none of these accounts can tell a wholly non-circular story about causality. That is, they are **non-reductive accounts of causality**. Next, these causal interactions between elements are **organised** in some way. Mechanisms are not simply aggregates of parts. Instead, the components of the mechanism have an internal structure that contributes to their eventual function. This suggests that giving a description of something other than the individual physical parts and their associated behaviours is a necessary part of describing a mechanism. Part of this claim is the suggestion that upstream causal effects of the mechanism have an effect in the later parts of the same mechanism. But to claim that this organisation is simply a linear causal sequence is mistaken. Instead, mechanisms may internally interact in complicated ways. So, for example, a downstream effect of a mechanism may act causally upstream in a way that is vital for the proper function of the mechanism in subsequent instantiations. So an account of organisation will take account, not only of the sequential choreography of elements and their activities, but will also have to tell a story about how the mechanism is chronologically located in context. It is not simply the case that a described mechanism runs once through, from start to termination, in isolation. Instead, **causal mechanisms are necessarily embedded in their environments**. This means that we might also have to give an account of how our mechanism is spatially, chronologically and contextually situated in order to give a usefully full account of it as a mechanistic model.

These accounts also feature some variant of **regularity** or **stability**. Mechanisms are mechanisms just because they are stable, to some degree and in certain contexts. This stability arises from the causal relationships between elements. In other words, mechanisms arise because of organised **causal dependencies**. Each element, with the possible exception of those at the extreme 'left' are brought about by the effects of their causal parents. This does not mean that a given mechanism must always turn out the same group of effects. In what follows, I try to unpick the nature of this stability. Can we give an account of what makes mechanisms behave regularly, and what prevents them from doing so? This appears a key issue in an operational theory of causality for medicine. After all, if we can't suggest why a person becomes unwell, as a mechanistic consequence of the disruption of their usual (mechanistic) physiology,

[13] There have been a number of previous attempts to synthesise various mechanistic accounts of causality. See, e.g. Tabery, 2004.

then this project is in trouble. I try and pick out the manner in which regularity becomes irregular in section 5.6.

This aspect of (ir)regularity raises a further consensus item regarding a distinction between mechanisms and mechanistic models. Mechanistic models need to be **functionally specified** if they are to give a meaningful mechanistic description of a phenomena. The heart is (part of) the mechanism for pumping blood around the vascular tree. But it is also (part of) the mechanism for producing the heartbeat, for filling the middle part of the mediastinum, for occupying the pericardium, for supporting the coronary arteries, for regulating blood pressure, for secreting atrial natriuretic peptide, and for contributing to the development of a great many disease states. When we want to employ the heart as part of a mechanistic model, we must first specify what purpose our construction of the model is to serve. This functional specification will control exactly the nature of the entities and activities contained within the mechanism, what organisational concepts we employ, the nature of theories governing interactions between elements, and the types of chronological and spatial contextualisation we employ. The heart mechanistic model that we might use to understand why the anterior and posterior parts of the mediastinum do not collapse into one another will be very different from that heart model that we might use to understand the manner in which a particular form of cardiomyopathy has affected serial generations of a family. In turn, this will be a very different mechanistic model from that we might use to explaining why an individual has high blood pressure, should refrain from smoking or has pericarditis.

As can be seen from this outline of functional specification, I detect a further consensus regarding the **interlevel nature of mechanisms**. Mechanisms are inherently interlevel. There appears to be no single (ontically, epistemically, methodologically determined) privileged level of organisation at which the mechanism can be said to exist, to be understood or should be investigated. In other words, mechanisms give a *non-reductionistic account of causality*. Take the above sketch. The heart, as a mechanism, can be understood at a variety of levels of organisation. We can give an account of the heart as a anatomical tissue,[14] or by the behaviour of its component cells, or by the mechanistic properties of these cells, or by some much more detailed story of its biochemical, molecular biological, genomic, chemical, physical aspects, and so on. Yet a properly, functionally specified mechanistic model of a particular heart-y behaviour, such as "produces heart sounds" will require elements of causal description at many of these levels without specially privileging any of them. While a biochemical explanation of, say, myosin-actin interactions will be required to understand why the heart beats, the appropriate level of organisation for understanding the way that the chambers of the heart actually contract will require telling a tissue-level story about their spatial arrangement. So giving a mechanistic account of something requires relating together the manner at which it functions at a number of different levels simultaneously. I will discuss this at greater length in 5.4.

These accounts of mechanism are also **metascientific**. That is, in general they base themselves firmly on the study of scientific practices, rather than beginning from a metaphysical stance on causality. Usefully for my account, this scientific starting-point has generally been a biomedical one. Most com-

[14] A structured organisation of cell-types. Incidentally, I think that terms such as this require at least a partial mechanistic understanding in order to play a useful semantic role. But this digression is for another time.

monly, the specific disciplines employed have been neuroscience (Glennan, MDC, Craver) and molecular biology (Bechtel and Darden).[15] Perhaps as a consequence (or cause?), these theories have some degree of **normative aspiration**. They do not simply aim to describe causes.[16] Instead, they aim to provide a cognitive tool for determining what are good, and what are not so good, causes in the world. I will pick up some of these normative aspirations in chapter 7.

Finally, and closely related to the normative component of these theories, all of them in some way distinguish the mechanism in the world from our models of these mechanisms. They also distinguish these models from our representations of them. Entities, activities, various types of temporal and spatial organisation and their resulting causal dependencies found in mechanisms are real, ontological features of the world. However, this does not mean that the new mechanistic philosophy makes claims solely about the ontology of causal mechanisms; quite the opposite. Instead, there are a number of consensus features - notably about the interlevel nature of mechanistic models, and the various discussions about how to specify these models - that are claims about how mechanisms should be modelled. That is, these are epistemic claims. Finally, there are also a group of metaphysical claims about mechanisms and their models, most prominently those claims that deal with the metascientific nature of the mechanisms project, and various normative claims. However, in what follows, I will primarily engage epistemically. While I claim that entities, activities, causal dependencies and so on are real, mind-independent features of the world, I think that the overarching problem of dealing with causes in practice concerns issues related to giving accurate models of these phenomena. So in what comes after I will neglect the ontic aspect of mechanisms. Instead, the question will be: what do we think is the cause of a phenomena? Why do we think this is so? How should we model the causes of a phenomena? It is this sort of reasoning about mechanistic models that I wish to use as part of my account of causality in medicine. To give an historically adequate account of such mechanistic modelling, I will need to give a more particular account of what I think constitutes an adequate model of a mechanism. So I will now break away from this consensus position on mechanisms in order to achieve this. I have a clear preference for one particular mechanistic theory of causality for the job in hand, and I will adopt a more partisan stance below.

5.3 Mechanisms in medicine

These competing accounts of mechanisms have major differences when it comes to making mechanistic models of medical phenomena. While each of them have something to add, MDC's account of mechanistic causality appears to be the most suitable account upon which to build the expanded RWT. Why MDC?

Following MDC's account, I argue that we need to consider both activities and entities when we model causal relationships in medicine. There are sound methodological and epistemic reasons for making this claim: both types of thing are, I suggest, necessary components of making good mechanistic

[15] Woodward is an obvious exception to this aspect of these accounts.

[16] Really causal explanations, in all cases. I have rather neglected the explanatory capacities of causal mechanisms in my account in favour of using them as tools for dealing with causality.

models. I will defend this claim below by showing that mechanistic models that neglect the role of either entities or activities are inadequate.

I also suggest that neither activities nor entities are prior. That is, I do not think that it is possible to reduce all talk about entities to being about activities, and nor is it possible to reduce all activity-talk to entity-talk. This is part of the more general sense in which this account is non-reductivistic, and non-reductionistic. Again, I will support this claim by reference to an historically-grounded argument in section 5.3.2. One important consequence of this argument, though, is that I do not agree with MDC about the type of dualism that obtains. While MDC make similar arguments from epistemology to support their ontic activity-entity dualism (Machamer, 2004), I will not extend my argument in such a way. In fact, the argument presented in section 5.3.2 seems to undermine MDC's attempt to make an epistemically supported ontological argument about dualism.

5.3.1 Activities without entities; entities without activities

First, a conceptual claim about medicine - and perhaps science - in general. We are very often aware of causal activities without knowledge of the relevant entities that produce them. Take the example of Lucké's clinical descriptions of renal adenocarcinoma in the northern Leopard frog (Lucké, 1934, 1938a,b, 1939; Schlumberger and Lucké, 1949; Lucké, 1952. See also section 1.1). Here, we have an example where an activity was clearly discernible, could be very clearly described, and appeared to play a central causal role in a plethora of mechanisms. Yet Lucké could not describe the entity(ies) that produced these activities in such a clear way. In a similar way, he also did not know just what entities this unknown entity had its causal activities on. In both cases, he could tell a partial story about the likelihood of these features arising from an infectious agent, and perhaps give some kind of general mechanism of oncogenesis. But what he could not provide was a detailed mechanistic model relating the clinical findings to a particular entity. Thus, the mechanistic model for this condition would be inadequate. We can have a degree of causal knowledge (as Lucké did), without having a good mechanistic model.

The converse position, where we have knowledge of an entity without knowledge of a corresponding activity, might be illustrated as follows. Imagine the example of a discovery of a viral protein of unknown function. This constitutes an entity that lacks a clear activity. Yet it may well be the case that this entity appears to participate in a causal mechanism. For example, we might know about its structure, we might give an account of how it comes about or we might also be able to say that it causally participates in some manner downstream. For instance, we might conduct an experiment where a mutant virus, which does not express the particular protein, was used to infect cells in culture. We might then be able to give an account of what happens, and relate it in a contrastive way to the operation of the wild type virus on similar cells. But all of this does not mean that we know about the activities of this protein in detail. Activities show how entities act as causal agents; they are the way that they 'express themselves' (Machamer, 2004, 29). But this does not mean that entities without known activities do not 'express themselves'. In fact, as experiments such as this mutant virus show, we may very well know *that* a particular entity expresses itself in a mechanism, and may be able to give a clear account of *what* form this expression takes. What we do not know is *how* this expression happens. In RWT terms, we can

show that infecting culture cells with a mutant virus makes a difference compared to the effects of using wild-type virus. This tells us about the existence and likely shape of the underlying causal mechanism, but does not grant us detailed mechanistic knowledge about how the two situations differ. I will return to this distinction between types of causal knowledge later in this chapter.

In summary, such unilateral knowledge of either activities or entities does not allow us to adequately understand the relevant causal mechanism. Thus, knowledge about both entities and activities is required to construct a sound mechanistic model. This requirement is highly significant for the use of mechanisms within the RWT, because if we do not understand how a mechanism operates - to some extent - then it cannot become integrated with other forms of evidence, and thus it cannot meaningfully participate in supporting a causal claim. Knowledge about both activities and entities is necessary to formulate an adequate causal mechanism.

5.3.2 Some entities are really activities; some activities are really entities

A different historical argument supports the position that neither entities nor activities are prior. This argument depends on demonstrating that, dependent on the context, some activities may be treated as if they were entities, and some entities may be treated as activities.

I will begin by illustrating just how this *interconversion thesis* might manifest itself. For example, some (medical) entities are defined in terms of activities. The term *extracellular space* is defined as the area surrounding a cell which is the venue for local cell-cell interactions. But when we employ this entity in a mechanistic model by, say, speaking of interactions that occur in it, this entity is not characterised in terms that we generally associate with entities. We do not give a definition in terms of spatial characteristics, morphological features, biochemical characteristics and so on. Instead, this area is defined by a group of local interactive activities which can be (partially) described in componential terms. Similarly, some activities are defined in terms of entities and their downstream activities. So the activity *T-cell activation* is defined in terms of the entities that undergo activation. We cannot, in this case, tell a fuller story about what the activity *T-cell activation* entails in terms of componential activities and entities.

I take these kinds of cases to suggest that, in certain contexts, we may - methodologically and epistemically - treat one kind of thing as if it were the other. However, we still require both kinds of thing if we are to make sound causal claims - as discussed above, claims that lack either activities or entities tend to be unsound. However, the context-dependent nature of activities and entities suggested by the interconversion thesis means that we should not expect the (epistemic and methodological) dualism that is required to formulate sound causal claims to be necessarily underwritten by a corresponding ontological dualism.

5.3.3 A rebuttal of arguments against activity-entity dualism

Now there are a number of possible objections to this group of arguments in favour of dualism. First, perhaps the Lucké case represents simple historical contingency. We happen to have discovered the clinical syndrome well in advance of its related causal entity. But this chronological consideration does not licence any argument about the relative statuses of entities and activities. I suggest that this sort

of argument is no licence at all to discard dualism. Imagine the counterfactual case, where the clinical syndrome and its associated viral entity were discovered simultaneously. In this case, many (functionally specified) causal mechanisms would be unchanged. We would still have a similar understanding of the sorts of downstream causal interactions of the tumour syndrome that we do now. So our mechanisms dealing with frog survival and so on would be - in almost all contexts - unchanged. What would change would be the upstream causes of the syndrome. Detailed knowledge of the relevant virus would (as with the HBV case) allow us to intervene, make better predictions and so on, and it is this kind of difference that makes the mechanistic model an adequate one for medical purposes.

A second, related, counterargument: perhaps activities are simply the capacities, propensities or properties of an entity? Perhaps, in this case, the knowledge that kidney tumours and RaHV-1 are related in certain law-like ways is sufficient. As a further example, aspirin can relieve headache. But it can also prevent arterial thrombosis and cause peptic ulcers. We can now give a (molecular) account of all three of these activities, and relate them to a single property of aspirin. Aspirin inhibits cyclooxygenase. Thus, it blocks pain pathways - relieving headache. It prevents platelet aggregation - preventing thrombosis. And it also causes gastric erosions, leading to gastrointestinal bleeding. In this case, the property of aspirin to relieve headache (and pain more generally) was described well in advance of the other two properties. In no small part, this was due to the simplicity, the obvious Humean constant conjunction, of cause and effect. The other properties are rather more difficult to understand, because their effects are temporally extended, may require various technological developments to detect, act in a stochastic fashion, interact with manifold other causal variables and so on. But arguments of this sort un-nerve me. Consider the effects of opium. Opium was described as having a dormative virtue that was the root of its narcoleptic powers.[17] But this is simply begging the question: users of opium sleep, because opium has the capacity to evoke sleep. In general, like the RaHV-1 case, I think we need to tell a fuller story of the relationship between activities and entities, rather than one that is so tightly circular. This means that we must give a full account of both entities and activities to adequately describe causes mechanistically.

5.3.4 Against other mechanistic theories of causality

So even though we disagree about the ontological implications, dualism is the main positive reason that I prefer MDC's account. There are also important arguments that count against the other mechanistic theories of causality. Very briefly, Glennan I (Glennan, 1996) seems untenable because of its requirement for causal laws. In biomedicine, even the use of the term appears lacking. Take the cases discussed in chapters 2 and 3. I can find no pertinent reference to the suspicion of, search for, discovery or evaluation of any form of causal laws governing any actual or theorised causal interactions anywhere in the (extensive) literatures on these cases. This lack of support for even the possibility of detecting causal laws seems to me sufficient to discard Glennan's first account, on the grounds that the metascientific case lends no support to the metaphysical heart of the theory. That is not to say that everything Glennan discusses is incorrect. In fact, I modify and employ several aspects of his account - most notably his dis-

[17] '*Opium facit dormire... Quiat est in eo / Vertus dormitiva, / Cujus eat natura / Sensus assoupire.*' [Opium brings sleep because in it there is a dormative virtue that causes drowsing of the senses] (de Moliaure, 1673)

tinction between mechanisms and models of mechanisms, and his account of mechanistic polymorphism - later.

In a similar fashion, Woodward's interventionist account lacks historical support in the historical cases. In fact, many pieces of evidence that actually played a crucial role in suggesting, supporting and disproving causal relationships appear very difficult to reconcile with this manipulationist account. For example, the prospective study that played such an important role in the BL story seems to violate many of Woodward's conditions on interventions.[18] Likewise, Burkitt's initial clinical work that formed the foundation for the BL research programme seems vital, but cannot be intelligible under a theory that speaks only to interventions. But this is not to say that interventions played no significant role in causal discovery and assessment in the case histories. In fact, the opposite; they played a critical role. It is just that thinking in terms of interventions alone does not capture the actual richness and plurality of evidence in play in building and assessing causal claims in practice. Thus Woodwardian interventions appear a necessary, but insufficient, component of constructing mechanisms. It therefore seems improper to base my account on them, instead preferring a more general account of mechanism that can accommodate a greater range of types of evidence.

There is a second way in which Woodward's account fails to account for the historical case. This is in Woodward's treatments of **modularity**. Modularity is a property of representations of causal systems,[19] which he defines as:

> **Modularity**. A system of equations is *modular* if (i) each equation is level-invariant under some range of interventions and (ii) for each equation there is a possible intervention on the dependent variable that changes only that equation while the other equations in the system remain unchanged and level-invariant.
>
> (Woodward, 2003, 329)

The important consequence of modularity is that, if a system is modular, then we can perform certain interventions on single components of it that we can subsequently expect to affect only that single component. That is, assuming modularity means that we will be capable of performing specific, 'surgical' interventions within a given causal system. Woodward illustrates this property with the familiar example of the causal relationships between atmospheric pressure, a barometer and the occurrence - or otherwise - of a storm (Woodward, 2003, 48ff).

Let us model the relationships between the reading on a barometer (B), the probability of a storm occurring (S), and atmospheric pressure (A) by the following pair of equations:

(1) B=$a$$A$; and

[18] Although the divorce of interventions from human agency seems a valuable feature of Woodward's account. For instance, the geographical distribution of BL and its relationship to climatic factors is, essentially, a natural experiment.

[19] Woodward claims that it modularity is a necessary feature of an adequate representation of a causal systems as either a system of equations or as a directed graph: 'I assume that when causal relationships are correctly and fully represented by systems of equations, each equation will correspond to a distinct causal mechanism and that the equation system will be modular.' (Woodward, 2003, 49)

(2) $S=bA$

If we assume that our representation is modular,[20] then we expect to be able to intervene on (1) by replacing $B=aA$ with $B=I$ - by, say, forcing the needle of the barometer to a particular position and nailing it into place. In a correctly modular representation, such as this one, an intervention within one equation will not affect any of the others. Here, intervening on the barometer will not alter the chances of a storm taking place. Likewise, if we were to make $S=I$, by performing some intervention to alter the probability of a storm occurring, then assuming that this intervention did not rely on altering A, B would similarly remain unchanged.

Here is the difficulty with assuming that modularity must hold in order to provide an adequate representation of a causal system: as modularity is an empirically determined property of representations of causal systems,[21] it may be that we have a causal model that is quite adequate for describing particular aspects of the system of interest, and yet in which modularity does not generally hold.

As an example, take the role of statins in preventing coronary heart disease. While the intended effect of these agents was to inhibit cholesterol synthesis, and thus prevent heart disease by decreasing circulating levels of lipoproteins, it transpires that they have additional effects that are independent of their cholesterol-lowering role.[22] In fact, statins also appear to prevent heart disease by 'improving endothelial function, enhancing the stability of atherosclerotic plaques, decreasing oxidative stress and inflammation, and inhibiting the thrombogenic response' (Liao and Laufs, 2005, 89). Further, these pleiotropic effects have been seen to contribute to survival in a number of recent clinical trials (Wang et al., 2008).

Now these pleiotropic effects potentially compromise modularity in any system that attempts to model the causal relationship that obtains between cardiovascular outcomes cholesterol by intervention with lipid-lowering statins. For not only will the statins alter lipid levels, but they will also affect other contributors to coronary heart disease, such as inflammation, plaque stability, and so on. More importantly, this kind of pleiotropic behaviour is not unusual in causal systems of medical interest. In fact, it has been found in a great many systems, including with anti-hypertensive drugs (Koh et al., 2003), in the effect of genetic factors (Flanagan et al., 2000), in the action of immune cytokines (Veldhuis et al., 1991) and so on.

So if we are to take seriously Woodward's requirement that adequate causal representations be modular, we must regard our model of the action of statins - and possibly a great many other mechanisms - as extremely suspect, even if we lack a better alternative.[23]

---

[20] It need not necessarily be so,'Modularity is thus a feature that a set of *representations* of causal relationships... may (or may not) possess.' (Woodward, 2003, 47).

[21] Woodward has previously suggested that modularity is empirically determined: 'We are not claiming that one can determine whether a system of equations is modular simply from its syntactic form... it is nature and in particular facts about what happens or would happen under interventions that determine whether a given system of equations is modular.' (Hausman and Woodward, 1999, 543-4).

[22] i.e. statins are pleiotropic.

[23] 'If we make the... plausible assumption that a necessary condition for two mechanisms to be distinct is that it be possible (in principle) to interfere with the operation of one without interfering with the operation of the other and vice versa, we have a justi-

This difficulties with Woodward's account also propagates to Glennan II (Glennan, 2002; Glennan, 2005). Despite this difficulty, many features of these accounts are useful to me, and I pick them up in subsequent sections. In particular, I will return to a more Woodwardian/Glennanian view in chapter 6, when I deal with the relationship of evidence to causality, and again in section 7.1, when I discuss effective strategies for finding therapeutic interventions.

Bechtel claims that the parts of a mechanism are primary, just because the operations that these parts have on one another cannot happen in the absence of their parent parts. In this case, the arguments made in support of MDC's activity-entity dualism serve as effective arguments against this account. Despite this, Bechtel's account is of great interest when it comes to the business of mechanism construction, and in the relationship between mechanisms and mechanistic models.

So that is why I prefer MDC. As will become apparent, I think that MDC's account needs some modification to cope with the historical cases. Now, though, I will turn to why, in general, I prefer these interlevel accounts of mechanisms to those that rely on either reducing causal- to non-causal processes, or those that relate causation to the nature of physical processes.

5.3.5 In support of a non-reductivist account of causality

Does accepting the slogan 'no causes in, no causes out' (Cartwright, 1994) mean that this analysis of causality is, ultimately, pointless? I suggest that it does not. While theories of causality that do not reduce causal- to non-causal talk are circular, they are not necessarily viciously so. A well-construed, non-reductivist, circular theory of causality can be enormously helpful in understanding causality in practice. All of these advantages result from the leverage that such a theory allows us to exert on complicated causal situations. In other words, our theory is useful because it allows us to cognitively digest large, complex causes into smaller, more simple ones that are more intuitively understandable. Therefore, I feel no sense of queasiness in simply treating causation as primitive when doing so allows us to understand causes of practical importance.

In particular, I hope this theory is useful in the following ways: First, as a conceptual tool for understanding complex causal structures. Second, to allow us to develop effective strategies for modelling and representing these complex causal structures. Third, to allow these models and representations to reflexively assist in causal understanding by analogy and dissimilarity. That is to seek the ways that apparently disparate causal situations can be unified, or to distinguish causal situations that appear deceptively similar. Fourth, these models and representations may allow us to see how we can improve our knowledge of these causes by showing us effective strategies for mechanistic construction and evaluation - possibly by interventions. Fifth, perhaps related, they can grant us a greater understanding of the nature of the relationship between our evidence for our cause and our cause itself. Especially importantly in this case, they help us plan and operate effective therapeutic interventions. Finally, they provide an important tool for understanding the way that understanding of causes develop. As Machamer puts it:

> Just as one cannot have, or does not need, a theory of organism per se and *tout court*,

fication for requiring that systems of equations that fully and correctly represent causal structure should be modular.' (Woodward, 2003, 48).

> equally one does not need a theory of *cause*. The problem of causes is not to find a general and adequate ontological or stipulative definition, but a problem of finding out, in any given case, what are the possible, plausible and actual causes at work in any given mechanism.
>
> (Machamer, 2004, 27-8)

Just as biologists do not have a theory of organism to understand organisms, philosophers do not need a theory of causality to understand causes.

5.4 Levels and gaps

Useful models of medical causes are generally interlevel. That is, their explicatory power relies on relating phenomena at different levels of organisation together. For example, a given model might connect the molecular causes of a disease with a range of clinical symptoms and signs that occur at the level of the organism. In turn, antecedents of the molecular causes of the disease might consist of a range of factors at differing levels of organisation, and these might operate via intermediates at different levels.

In this section, I will begin by introducing two reductionistic accounts of causality. I will then attack these theories on largely historical grounds. This attack is intended to demonstrate that, contra these accounts, there appears to be no privileged level at which causal relata must operate. I will then make a number of positive suggestions in defence of an interlevel account of causality, before outlining two areas of interlevel accounts of causality that are in need of some development.

5.4.1 Process theories of causality

Salmon's theory of MT attempts to give a reductivist definition of causality that makes causality an objective feature of the world. For Salmon, to explain something is to show how it is situated within the causal nexus of the world. His account begins with Reichenbach's account of the *Mark Principle* (Reichenbach, 1956, 197ff), in which Reichenbach suggested that the transmission of marks could be used to determine causal relevance relations between events.

Salmon's development of this theory, in Salmon, 1980 and Salmon, 1984, was as follows. He first claimed that processes, rather than events, were the basic entities of causality (Salmon, 1980, 50). The connection between cause and effect could then be interpreted in terms of the workings of physical processes (Salmon, 1998, 16). However, because only some physical processes are truly causal ones, this development required Salmon to suggest a means of differentiating causal processes from non-causal pseudo-processes. Salmon suggested that we can make this differentiation by determining the capability of a process to transmit a mark (Salmon, 1980, 52). Causal intersections are those that transmit persistent marks when they interact in Minkowski spacetime:

> We can distinguish two types of intersections: causal *interactions* and noncausal *intersections*. When there is an intersection between two processes in which both are modified, and the modifications persist beyond the place of intersection, this intersection qualifies as a causal interaction. One uses a causal interaction to produce a mark in a process.
>
> (Salmon, 1998, 17)

So causal processes are those that transmit marks, and marks are transmitted by causal interactions (see also a more technical definition in Salmon, 1984, 139ff). Salmon therefore claimed that MT could give an adequate, objective, characterisation of causal interactions in non-causal terms. However, Kitcher (1989) suggested that these claims were mistaken. His first line of attack was to suggest that MT could not give an adequate analysis of causality. Kitcher identified three general lines of criticism against the adequancy of the MT theory (reviewed Kitcher, 1989, 469). First, Kitcher claims, the practical differentiation between causal processes and non-causal pseudo-processes is very much more complicated than simply determining whether or not they transmit marks. In fact, many apparently mark transmitting processes do not, when examined carefully, actually transmit a mark at all. Instead, a range of situations exist where: '...the existence of unwanted (spurious?) characteristics - pseudomarks, pseudomodifications, and the like' confounds this distinction (Kitcher, 1989, 469). So Salmon's - supposedly simple - differentiation between types of processes by MT is practically very complicated. Second, and likewise suggesting that this practical differentiation is unworkable, the inertial properties of some (causal) processes may mean that we can apparently mark them with pseudoprocesses. Third, this account relies upon reducing causality to the operations of idealised, elementary processes. This means that we have to account for the behaviour of the macroworld in terms of series of these elementary, idealised processes. However, this presents us with a methodological difficulty, in that it makes our ordinary causal knowledge parasitic upon more fundamental knowledge of these idealised processes. Kitcher argued that this was simply not the case.

In addition to these criticisms on the grounds of adequacy, Kitcher also suggested that the MT theory was simply a species of counterfactual theory of causality.

> I shall try to show first that the counterfactual commitments of Salmon's theory are far more extensive than he has noted (indeed, I shall claim that the theory naturally evolves into a counterfactual theory of causation), and second that the epistemology for counterfactuals that Salmon outlines encounters grave difficulties.
>
> (Kitcher, 1989, 470)

It is the case that counterfactual conditions are really required for this account to give an adequate characterisation of causal processes. Defining causal processes in terms of transmitting persistent marks actually requires a characterisation in terms of the *capacity* of these processes to transmit marks (Salmon, 1998, 19). In other words, assessing whether or not a process transmits a mark or not requires the interpretation of a counterfactual statement regarding what would have otherwise happened.

Another, even more penetrating, critique of the MT theory came from Dowe (Dowe, 1992; developed further in Dowe, 2000). Dowe suggested that Salmon's MT theory was flawed in four respects. First, it was circular, in that the definition of causal interactions in terms of mark-production begged the question (Dowe, 1992, 200-1). Next, that Salmon's definitions of MT were vague (Dowe, 1992, 201-4), and therefore could not really be used to differentiate causal from non-causal processes. Dowe then suggested, on statistical grounds, that two important types of causal interactions, Y- and λ-type interac-

tions, could not be deemed causal by MT (Dowe, 1992, 204-7). Finally, in common with Kitcher, he demonstrated the substantially counterfactual nature of the the theory (Dowe, 1992, 207ff).

However, unlike Kitcher, Dowe then suggested a possible modification to Salmon's theory that could save it from this range of objections. Instead of characterising causal process in terms of MT, Dowe suggested a definition of both causal interactions and causal processes in terms of the manifestation of *conserved quantities*, which he defined as:

> ...any quantity universally conserved according to current scientific theories. Some conserved quantities are mass-energy, linear momentum, angular momentum, and charge.
>
> (Dowe, 1992, 210)

So, according to Dowe, intersections are causal interactions if there is a CQ-exchange between the participating processes, while causal processes are characterised by the manifestation of CQ:

> DEFINITION 1. A causal interaction is an intersection of world lines which involves exchange of a conserved quantity.
> DEFINITION 2. A causal process is a world line of an object which manifests a conserved quantity.
>
> (Dowe, 1992, 210)

As with Salmon's MT interactions, these world lines are collections of points in Minkowski space-time (Dowe, 1992, 210). Dowe claimed that this characterisation solved all four of his objections against MT. First, defining causal processes in terms of conserved quantities was non-circular, as it did not rely on the inherently causal notion of marking. Second, this formulation was much simpler that Salmon's (rather complicated) later formulations of the MT theory, and was thus not vague (Dowe, 1992, 213). Again, unlike MT, the CQ account could correctly interpret a wider range of types of intersection, notably including Y- and λ-type interactions (Dowe, 1992, 214). Finally, this modification also freed this version of the process theory from a reliance on counterfactuals, and made it capable of giving a fully objective account of causality.

Salmon was to admit that Kitcher and Dowe's criticisms were correct, and that they meant that MT could not give a fully objective account of causality (Salmon, 1984, 171ff).[24] He therefore adopted a modified version of Dowe's CQ theory, such that it is the transmission (rather than manifestation) of CQ that is the signifier of causal processes:

> the crucial question regarding causal processes is what they do on their own without outside intervention. My answer is that they transmit something-e.g., conserved quantities, information, or causal influence-and it is by virtue of such transmission that events at A and B ... are causally related. Dowe gives a different answer.
>
> (Salmon, 1997, 466)

[24] This debate led to a fascinating exchange of opinion in the philosophical literature (Hitchcock, 1995; Dowe, 1995; (Salmon, 1997); also see Salmon's overview in the postscript to the revised version of Salmon, 1994 contained in Salmon, 1998, 260)

5.4.2 Against reductionist accounts of causality

First, there are a number of well-known objections against process theories (reviewed in e.g. Williamson, 2010). Broadly, these fall into two camps. First are a range of objections that arise from quantum mechanics (violations of the common cause principle, retrocausation and action at a distance). The second objection is a much broader one: reducing most phenomena to microphysical processes is practically problematic.

Now this second objection receives much support from the historical cases. First, the medical literature reveals no clear level of organisation which is assumed, or even argued, to be ontologically privileged when it comes to demonstrating causal relationships. It is not the case that debates around these causal claims were framed in terms of their microphysical characteristic. Further, there seems no support for any more limited, domain-specific reductionistic programme.[25] Again, it is simply not the case that arguments about causation are framed in terms of the knowledge of causes at one particular level. When it comes to causal claims, no one special level of organisation is historically apparent. A closely related argument can be used to gain further support for this claim: there is a paucity of serious attempts at reducing biomedical processes to lower-level phenomena. In the scarce instances where these attempts are successful, they are highly constrained to particular domains. For example, it may be exceptionally useful to give an explanation of biofilm adhesion in terms of quantum field events. But events at this level seem largely irrelevant when it comes to giving an explanation of common causes of falls in the elderly. Moreover, in the first case the microphysical explanation suggests likely means of intervening, but the second case does not.

Second, a study of clinical medicine reveals the centrality of thinking about causes in an interlevel fashion. For instance, by far the majority of diseases have manifestations at the level of the whole organism.[26] When a diagnosis is made, these whole-organism findings may be joined by the results of a range of investigations that examine the disease state at a wide range of organisational levels. These might include various imaging modalities, including X-rays, ultrasonography, CT or MRI scans, blood tests for ions (sodium, potassium), small molecules (glucose, urea), polypeptides and proteins including complexes of antibodies or enzymes, tests of blood or other bodily fluids for microorganisms, tissue biopsies or cytological sampling procedures to examine cells and so on. Finally, evidence may well arise from therapeutic interventions that pertains to causation. All these findings form the evidential basis for making and assessing causal claims. Thus claims about a privileged level, on methodological or epistemic grounds are false. As Bechtel puts it:

> Explaining a phenomenon is in part a matter of finding the correct level for understanding particular interactions. Moreover, the resulting model is inherently interlevel. Explaining how a system works involves not only determining how it interacts with other systems, but what its components are, what they do, and how they are integrated so as to enable the

[25] This is similar to the difficulties experienced with other, intertheoretical reductions in the biological sciences (Kitcher, 1984).

[26] We generally know these as symptoms and signs. Symptoms are those findings that are obtained from history-taking. That is, they are phenomena which the patient reports. Signs, on the other hand, are elicited by the examining clinician. A headache is a symptom, a swollen optic disc is a sign.

> system to perform the activity of interest. It should be emphasized that in developing such mechanistic explanations, scientists do not generate anything resembling a theory reduction. They do not develop two theories and connect them with deductions, but rather models that cross levels or interlevel theories
>
> (Bechtel, 1994, 16)

A range of more specific, empirical, objections might also be levelled against the MT and CQ theories. Many common biomedical phenomena actually do not appear to involve the exchange of a conserved quantity or the transmission of a mark. Take an extremely common type of cellular interaction - that of a ligand binding with a receptor. As the ligand (a hormone, for example) binds to the receptor it induces a conformation change in the receptor. In turn, this activates a range of various cellular machines. However, as far as I can see (and others too, such as Craver (2007, 75-7)) there is no interaction that is explicable in terms of CQ or MT that occur between the ligand and the receptor.

But these arguments seem less convincing than those originally raised against the MT theory by Kitcher. We simply do not, in medicine, make, appraise or defend causal claims by discussing the operations of physical processes. Our causal knowledge of the macro-physical world is not, in this case, dependent upon arguments about fundamental processes. This argument seems sufficient to licence the rejection of all three of the process theories (MT, and both variants of CQ) as unsuited for understanding causality in medicine.

5.4.3 In support of interlevel models of causality

So if an historical reading fails to support the claim that the only real causes are those that operate at a fundamental level, what is the alternative? I suggest that mechanistic models provide the means by which non-fundamental phenomena can participate in useful causal models. In fact, I suggest that most mechanistic models will only provide useful causal models if they deal with the relationships between non-fundamental phenomena. This is not a controversial claim, at least for the new mechanistic philosophers (see, e.g. Craver, 2005; Craver, 2007, 11-18). In fact, there may be grounds for arguing even more strongly than this. Some have suggested that this sort of interlevel aspect of mechanisms is what grants them any explanatory power at all:

> It is the integration of different levels into productive relations that renders the phenomenon intelligible...
>
> (Machamer et al., 2000, 23)

I think this suggestion is correct. However, I will need to introduce a few more concepts before I can clearly say why. I will therefore resume, by justifying this claim, in section 7.2. For now, I will deal with an important aspect of conceptualising different levels in practical mechanistic models.

So, unlike those accounts that attempt to reduce causality to fundamental physical processes, the causal models encountered in practice tend to bottom-out in domain-relative fundamental causal interactions. Essentially, within a domain of practice, these interactions are assumed to be causally primitive.

We can assume this because these kinds of interactions are likely to be the most striking, simple and immediate kinds of causal interaction experienced in the domain.

This bottoming-out level will be specified in terms of action-oriented and cognitive goals. For medicine, this bottom is generally at a macromolecular level. Below this, we find interventions and understanding hard to do. For instance, when we talk about action potentials, we can give a rather specific causal story about the behaviour of proteins and other large molecules (ion channels, receptors) but only a rather limited, general one about the micromolecules (ATP, ions). We aren't really interested in the details of their individual activities, just their broad trends. We, similarly, tend not to intervene on these small molecules, instead giving, say, depolarizing agents that act on cell-surface receptors, and so on. But this neglect of fundamental phenomena is not, to me, a major concern. Even in the case that there is no necessity or necessary connexion between these low-level events, their obscure operations still provide the means by which we can reason causally at higher levels. If the mechanism, and hence our models of it, behaves as if there were fundamental, necessary connexion, even if there is not, then we do not need to worry about the origins of the apparent connexions that we uncover. For the purposes for which the causal mechanism is employed, the existence or otherwise of causation at a fundamental level is irrelevant. That is, even if we do not find a necessary connexion at the fundamental level, mechanistic models may still provide a vital means of mapping the causal territory. And it is this mapping capability that forms the backbone of my remaining account of the role of mechanistic models in understanding causality in medicine.

5.4.4 Levels are context-dependent

The choice of phenomena that should be included in a mechanistic model is highly context-dependent. This means that there is no objectively necessary or correct level, or range of levels, at which to describe causal relata. I suggest that this subjective choice of phenomena will be guided, as Russo and Williamson suggest about causes in general, by both cognitive and action-oriented reasons.

As an example, let us compare the differences between mechanistic model of cervical cancer that might be found in different contexts. A general gynaecologist needs to know lots of clinical details about cervical cancer: she needs to be capable of using the model to underwrite diagnosis, treatment, decisions about prognosis, and so on. But she does not need to know a great deal about HPV itself. Apart from knowing, for example, some details about which types of the virus confer a high chance of developing cervical cancer, the manner in which the virus is transmitted, and how cervical cancer is prevented by HPV vaccination, she does not need to know anything about the mechanism by which the virus acts. Instead, her mechanistic model should allow her to understand what is happening (and what is likely to happen) to a particular patient with cervical cancer, and to attempt to either prevent, cure or ameliorate the disease.

On the other hand, a research virologist will need to causally understand the precise manner in which the virus acts on cervical cells if he is to understand the viral life-cycle, or to develop new kinds of specific interventions directed against it. Now this isn't simply a question of the virologist having a mechanistic model that is simply more detailed about molecular-level events than the gynaecologist.

For instance, he would be able to quite safely ignore lots of the clinical details about cervical cancer that form such an important part of the gynaecologist's model of the disease. The two models differ in respect of the degree of detail with which they describe all manner of phenomena, and they do so for cognitive and action-oriented reasons.

This means that the gynaecologist or the virologist will use their specific causal model of the disease quite differently from one another. As an illustration, attempting to account for clinical phenomena, such as a particular visual appearance of cervical dysplasia at colposcopy, in terms of the patterns of protein expression of HPV is not only a poor strategy for the clinician; it is conceptually incoherent. For one, the reductionist component of the causal model is temporally removed from the phenomena. Second, the same essential causal story may result in a wide variety of different clinical problems depending upon the set-up conditions. Knowledge of how HPV proteins are expressed will not lead the diagnostician to the correct conclusion about the phenomenological features of the cervical anomaly. Nor does it provide an exploitable feature suitable for intervention. Instead, a story about the lesion in terms of types of higher-level causal processes (such as the different appearances of types of cervical tumours) would be appropriate, because it would lead to cognitive and action-oriented benefits. This suggests that there will be a degree of incommensurability between the two mechanistic models, even in such closely related forms of practice. Dealing with this sort of contextualised story about mechanistic models will be key to my account of causality, and I will address it further below. For now, though, my slogan might be that accounts of how causes operate at a fundamental level may be not only redundant, but might miss the point of giving the account at all. The context under which the cause operates is key.

5.4.5 Levels and gaps

As the above suggests, I think that mechanistic models will tend to be gappy. In the case of the differing models of the gynaecologist and the virologist, these gaps may well arise because part of the causal process might be irrelevant for their practice. Thus there may well be gaps in the causal mechanism that are *elective*. Given a particular context within which these mechanistic models are useful, it is not the case that the right strategy is to include all known details about the operation of the mechanism in the world. Instead, salient causal way-points[27] are sought that grant the user of the model cognitive or action-oriented leverage over the kinds of causal scenarios that they encounter in their practice. These way-points, and considerations of their relationship to practice, then guide the construction of the mechanistic model. This employment of elective gappiness is particularly prominent in representations of mechanistic models that are used for teaching purposes.[28] There will very probably be gaps that arise

[27] 'orig. U.S., a stopping-place on a journey; also, (on an air journey) the computer-checked coordinates of each stage of a long flight' (Anonymous, 1989)

[28] I think it is uncontroversial to suggest that simple mechanisms are easier to teach and to learn than more complicated ones. However, even the sorts of simplified representations of mechanistic models found in medical teaching materials tend to be very complicated, and a significant amount of interpolation is required to properly interpret them. This interpolation generally consists of filling in missing steps in a way that is consistent with background knowledge. The degree of gappiness in representations will therefore vary, depending upon the intended audience and their likely background knowledge of the causal situation that the representation deals with. There may be cases, though, where particular causal mechanisms act in a way that is generally inconsistent with background knowledge. In such situations, representations of mechanistic models may well have to include

from other sources too. Given the apparent failure of attempts to reduce biomedicine to physics, it must certainly be the case that we cannot give a full description of the operations of disease in terms of fundamental physical phenomena. However, this difficulty in fully explicating causal mechanistic models may well obtain at other levels too. Simply put, unless we have complete causal knowledge about the operation of a mechanism, it seems likely that our models of that mechanism will feature gaps that arise from a lack of knowledge. We might term this *obligate* gappiness to reflect the fact that it is unavoidable, at a given state of causal knowledge anyway. While this sort of gappiness may seem troubling, particularly because it seems to threaten the idea of productive continuity, I do not think it really presents any particular problem other than those practical ones that arise from the missing knowledge itself. After all, the gaps affect the mechanistic model, rather than the mechanism. One way that we can prevent missing causal knowledge from threatening the utility of mechanistic models is by carefully constraining their scope. This we do by *functional definition*.

5.5 Functional definition

As I suggested earlier, I am broadly realist about ontological mechanisms. That is to say that I think that mechanisms in the world are composed of mind-independent things that causally interact with one another in a manner that does not necessarily require, say, human agency in order to come about. Thus I claim some sort of objective existence of mechanisms. But this is not to say that I claim the same thing for our models of mechanisms, or our representations of them. In fact, quite the opposite: I suggest that models and representations of mechanisms are context-dependent. As Glennan suggests, mechanisms are behaviourally polymorphous, meaning that a single (ontic) mechanism may display a number of different (epistemic) behaviours that we might model. Likewise, each behaviour may have, as its cause, a number of different mechanisms. Thus the specification of mechanistic models will have to specify just which behaviours, and which ontic mechanisms for the behaviour, they are going to employ if they are to be adequate.

I suggest that this issue means that we require a process of functional definition when we build mechanistic models. The question '*what is the mechanism for x*' is only answerable when we define which aspects of the ontological mechanism it is that concern us. More specifically, we need to define our mechanistic model in terms of a carefully contextualised concept of production. So our mechanistic model for x must (and can only legitimately) be construed as the model that describes how x causally comes about. The process of contextualising the model fairly tightly constrains the sort of range of background (and possibly unknown causal factors) over which the mechanism is thought to act. This, then, is the broad phenomena that accounts for the differences between causal models of the same disease in different disciplines (section 5.4.4). That section was intended to demonstrate *that* mechanistic models are context-dependent. This section is designed to show *how* this happens.[29]

In the case of models (or representations) of medical mechanisms intended for clinical use, this will generally require a consideration, not only of the specific pathological routes that come together to form

some information about just why general background knowledge is a poor guide to understanding the missing steps.

[29] I think that a similar functional specification occurs when it comes to making representations of these models, too.

a productive continuity in producing the disease state, but will also require consideration of the relevant physiological, diagnostic and interventional pathways as encountered in practice. As an example, a clinical mechanism for cervical cancer might feature a number of HPV-dependent mechanistic routes that lead to cervical cancer. For instance, we might take common routes as those that 'go' via infection with HPV-16, -18, -33 and -35 (in the UK, at least). But in addition to these routes, there will also be a number that do not include these entities, but that still lead to the disease state. As none of these HPV types is a necessary cause of cervical cancer, it will be the case that many instances of the disease come about without infection with one of these high-risk types. This means that a model mechanism for cervical cancer intended for clinical practice necessarily includes the consideration of other causal routes that eventually lead to cervical cancer. Similarly, as none of these types are sufficient to cause cervical cancer, our clinical mechanism will also need to include causal routes that do not end in cervical cancer. In fact, as I will go on to suggest, I think that mechanistic models of a particular disease state need to include, in order to be intelligible in clinical practice, some mechanistic consideration of the normal physiological processes that would otherwise obtain.[30]

Now this inclusion of physiology might well not apply in mechanistic models employed by research virologists. In other words, the domain of intended enquiry is likely to change the precise specification of the mechanism. So functional specification will be highly context dependent. The functional specification of a typical therapeutic mechanism for cervical cancer will be much thinner than a research mechanism. This process of specification will demand modelling phenomena at different levels of organisation, describing the relations between these levels differently, including or excluding measures of uncertainty, prioritising different parts of the mechanism and so on. Functional specification, as well as the earlier concerns about levels, suggest that it is meaningless to talk of the mechanism for cervical cancer *simpliciter*. Why? Put another way, why can we not simply construct the unified mechanistic model for a particular disease, produce a suitable representation of it, and refer to this canonical text when we need to think mechanistically?

Leaving aside the (very important but shallow) practical questions regarding our lack of knowledge of most disease states, there are a number of more profound issues that such a unified approach raises. For instance, if we were to try and construct such a meta-representation, it might very well be the case that different parts of the mechanism would be incommensurable, as I suggested in section 5.4.4. For example, general gynaecologists may rely on a model of cervical cancer that depends upon certain properties of HPV. Yet the properties of HPV in this general gynaecological model might well be very different from those found in a model of cervical cancer used by virologists. HPV might be a rather different entity[31] or might have different activities[32] in each of the two models. Finding out just which way we should functionally specify our mechanistic model in a particular context will require a suitably contextualised account of evidence of difference-making, and I will therefore give a fuller account of

[30] In medical education, learning physiology traditionally antecedes learning about pathology. Perhaps this is the reason that so many disease states are conceived in terms of deviation from the normal physiology. Or perhaps this is the cause of the organisation of medical school.

[31] For instance, in the discussed details of viral protein expression

[32] For instance, in the manner in which these viral proteins act on different cellular machinery

this in the final chapter.

For now, though, I will move on to develop my suggestion that mechanistic models tend to display a characteristic *branched* morphology. This structure is a possible means of understanding just how difference-making evidence helps us when we come to functionally specify a mechanism.

5.6 Branching and track-switching

5.6.1 Introduction

One very important aspect of mechanisms that has been previously neglected is their ability to represent the sorts of branching relationships that play such an important part in the cognitive and action-related roles of causal relationships. This has critical ramifications for the use of mechanistic models as a component of the expanded RWT. In outline, branching arises from the fact that some causal entities have a range of possible activities, and that expression of these activities may depend upon the causal antecedents of the entities.

Let us take the insulin-secreting β-cells of the endocrine pancreas as example entities for a study of this branching behaviour. They have, depending on the particular situation, a range of different activities. The first of these is to secrete insulin at a rate determined by the concentration of glucose in their afferent blood supply. Note that the rate of insulin secretion is not a simple dichotomous variable. Instead, the rate of insulin secretion appears to be a differential function of the momentary concentration of blood glucose within certain bounds. Thus this continuous variable is one set of activities of the islet β-cells. Put another way, it is not the case that these entities either act, or do not act. Rather, their activities (in this case, the rate of change of insulin secretion) are determined in a continuous way. However, depending on the particular functional specification of our mechanistic model, we may find it more convenient to discretise such continuous variables. So we might - fairly arbitrarily - divide the variables *blood glucose* and *rate of insulin secretion* into 'high' and 'low' classes. In this simple case, a high blood glucose concentration would instantiate the islet β-cell's insulin-secreting activity at a high rate, while a low blood glucose would produce the opposite effect. Thus these activities would be the normal set of activities produced by these entities. Thus, this is a simple example of mechanistic branching.

But these β-cells may also have other activities which they do not generally express. For instance, the cells may - given certain antecedent conditions - violate this proportional relationship between blood glucose concentration and rate of insulin secretion by secreting insulin at an inappropriately high or low rate. So, for example, if these cells were caused to form an insulinoma,[33] they would exhibit an extremely high rate of basal insulin secretion, irrespective of the glucose concentration in their blood supply. Or they might, for some reason, fail to produce sufficient insulin and thus cause diabetes. All of these activities are those of the islet β-cells. So what is the difference between them?

[33] A kind of functional tumour of the islet β-cells

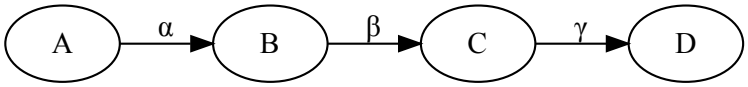

Figure 5.5: Simple physiological mechanism

5.7 Two kinds of effects

I think there is a clear differentiation between these two sorts of effect. The first, which I will refer to as a *sequence effect*, is achieved by the particular route taken through the mechanism. That is, this range of effects occur as a consequence of the mechanism operating normally, and are predictable given the normal set-up conditions that apply to it. In the example above, the range of effects on the rate of insulin secretion seen during normal homeostatic control of blood glucose would be instances of sequence effects. The second sort of effect, which we might call a *setting effect*, concerns a disruption of the way that the mechanism normally operate. Setting effects are different from sequence effects because setting effects alter the pattern of subsequent sequence effects. In the example above, the development of an insulinoma or diabetes would be setting effects.

As an even simpler illustration of the difference between sequence and setting effects, a sequence effect would be that sort of behaviour that we might find if we were to operate a toaster in the normal manner. This sort of pattern is characterised by consistency between different instantiations, in that we would expect a similar effect to obtain each time we insert the bread and depress the lever. We would not expect a given sequence effect to change the operation of subsequent instances of that effect. On the other hand, the effect of changing the settings on the toaster would be a setting effect. This would then have the effect of changing subsequent sequence effects - by, say, increasing the degree to which our bread becomes toasted. While both sorts of effect are intelligible under this conception of mechanism, it is setting effects that are really significant for branching. It is branching because of setting effects that shows us just how disease states are different from normal physiology. It is setting effects too that suggest the means by which we might effectively intervene in disease states. Think of a simple mechanism that might be used to represent a physiological causal process that results in the production of some sort of normal, physiological factor *D* (figure 5.5).

Say we then discover an individual in whom this mechanism appears disrupted - say, by them expressing *D'* instead of *D*. Here, we know that an individual has *D'*, but not how, or why, this obtains. Figure 5.6 represents such a situation, where an anomaly *(D')* is presented, alongside the normal physiology of this situation.

To find out how and why an individual has *D'*, we might then try to discover the mechanism that leads to *D'*-disease. We might find, in our single patient with *D'*-disease, that they also have two further anomalous entities *B'* and *C'* which appear to antecede *D'*. We might also find out about the activities

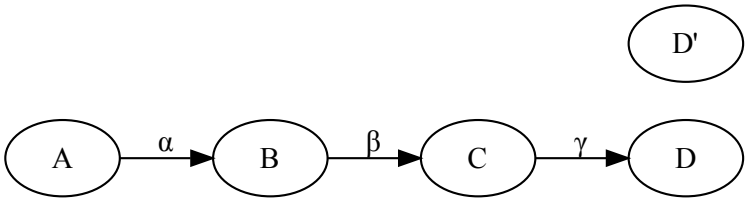

Figure 5.6: Mechanistic model of simple physiological mechanism with *D'* anomaly

of *B'* and *C'*. Thus we can construct a partial mechanism for *D'*-disease that relates it to *B'* and *C'* via activities β' and γ'. We can relate this partial mechanistic model for *D'* disease to the relevant physiological mechanism for *D* (figure 5.7).

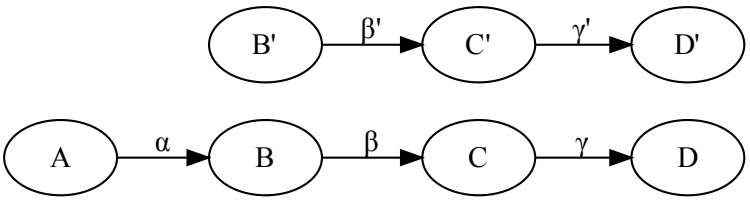

Figure 5.7: Mechanistic model of simple physiological mechanism, with partial mechanistic model for the *D'* anomaly

They then discover that *B'* arises from *A* (figure 5.8). This then gives a partial explanation of how *D'* comes about. However, in order to explain why this happens, we need to attempt to discover just why *A* has this divergent set of activities, α and α'.

This divergence of *A*'s activities is modelled experimentally, and is noted to be affected by the actions of three antecedent entities, *X*, *Y* and *Z*, and their associated activities on *A*, χ, υ and ζ. Certain combinations of these factors seem to cause *A* to do α, while others produce α'. Let us say that χ and υ causes *A* to α', while ζ causes α. In other words, χ and υ are promotors of α' at *A*, while ζ is a promotor of α (figure 5.9).

So far so good. Entities *X*, *Y* and *Z* act on *A* via their specific activities to produce two possible *A*-activities, α and α'. In turn, these activities cause either the normal physiological mechanism, or the pathological mechanism that produces *D'*. We can use this branched mechanism for a range of purposes. It is a knowledge of the setting effects on *A* that tell us how *D'* disease comes about. This knowledge provides us, first, with cognitive tools for understanding *D'*-disease. We can understand what is likely to happen in other cases of possible *D'*-disease. We can also employ it to develop an intervention, *I*, against

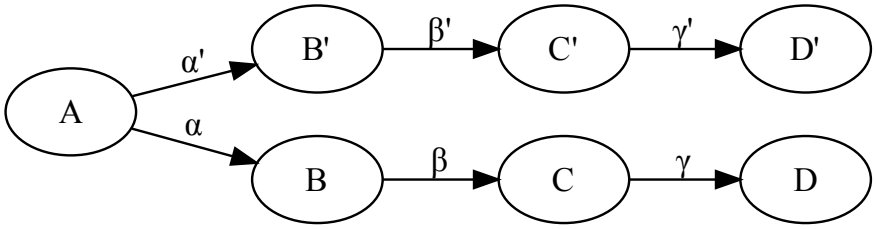

Figure 5.8: Mechanistic model relating *D'* to divergence at *A*

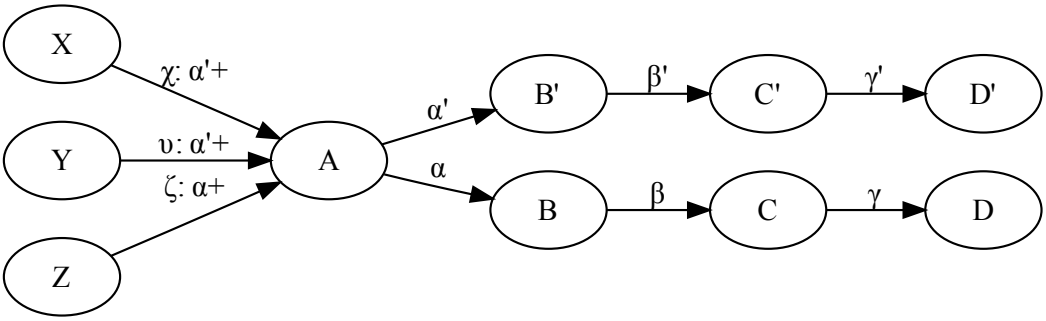

Figure 5.9: *X*, *Y* and *Z* as causes of divergence at *A*

D'-disease. We might, for example, choose to prevent *X* or *Y* becoming instantiated, thus preventing their χ-ing or υ-ing *A* (figure 5.10). This blocks the α*'*-ing of *A* and thus prevents the instantiation of the *B'* - β*'* - ... - *D'* mechanism.

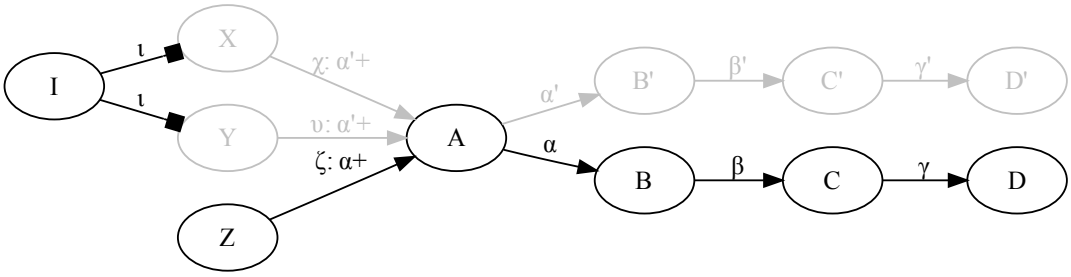

Figure 5.10: Preventing *D'* by preventing *X* and *Y* by intervention *I*

But this example hints at important non-mechanistic considerations that are required to interpret branching. How, for instance, do we know the effects of *X*. *Y* and *Z* on *A* in the two populations, *D* and *D'*? I will therefore return to this issue in chapter 7. For now, though, I will consider the role of branching behaviours in rather special circumstances: that of explaining causation by omission.

5.8 Preventions and omissions

The objections to negative causation of some kind are well worn (see, e.g. Aronson, 1971; Fair, 1979; Lewis, 1986, 189ff; Dowe, 2001). The essential problem is this: how can a cause that is a negative event instantiate an existent event? Take the following example, which is due to Beebee:

> Flora normally gratuitously waters her neighbor's orchids. But she stops watering them, and they die. Common sense judges that Flora's omission was a cause of the orchids' death, since had she watered them, as she usually does, the orchids would not have died.
>
> (Beebee, 2004, 294)

However:

> Common sense intuition identifies Flora's failure to water the orchids as a cause of their death, but not the failure of the other neighbors, and certainly not the failure of people on the other side of the world who neither know nor care about the orchids - even though it is perfectly true of each of them that had they watered the orchids, they would not have died.
>
> (Beebee, 2004, 295)

This issue affects many theories of causality, including counterfactual theories, processes and regularity. Mechanistic theories are no exception, as Craver suggests:

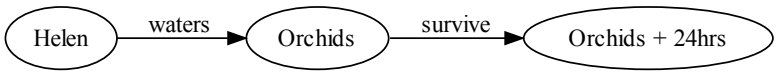

Figure 5.11: Negative causation example 1

Figure 5.12: Negative causation example 2

> ...in cases of omissions and prevention...there is no hidden connection between the cause and the effect. Such causes work by absences and gaps in connections...
>
> (Craver, 2007, 64)

If we admit one absence as the cause of a particular event, then we appear forced to admit a whole range of similar absences as causes also. This appears intuitively unappealing. Can we clarify the situation by using a representation of the relevant mechanistic model for this situation? I do not think we can. Let us consider a mechanism for Beebee's example (figure 5.11).

This represents the happy default state. Helen waters the orchids, which survive to bloom another day. So far, so good. What happens when she fails to water them is outlined in figure 5.12. Here, the orchids die because of Helen's failure to water them. But they also die because of the failure of any other individuals (for instance, Alfred, Bruce, Cleopatra, Davina and so on) to water them either (figure 5.13).

But these individuals seem, somehow, to be less causally responsible for the death of the orchid than Helen. In this example, Helen customarily watered these plants, while the other individuals did not. Thus there is some violation of normality that seems to make Helen responsible, and not the others. As Beebee admits, there are a number of similar contrastive relations that seem to act similarly in other cases of, intuitively appealing, negative causation:

> So it seems that common sense singles out an absence as a cause when, and only when, it stands out in some way - either from what normally happens, or from some norm of which the absence (generally an omission) counts as a violation. The norm might be moral, legal or epistemic...but other sorts of norm may well play a similar role.
>
> (Beebee, 2004, 296)

So far so good. Thus we may be able to save these sorts of problems by reference to some sort of norm. But this solution raises a further problem, in that it makes the analysis of causal claims critically

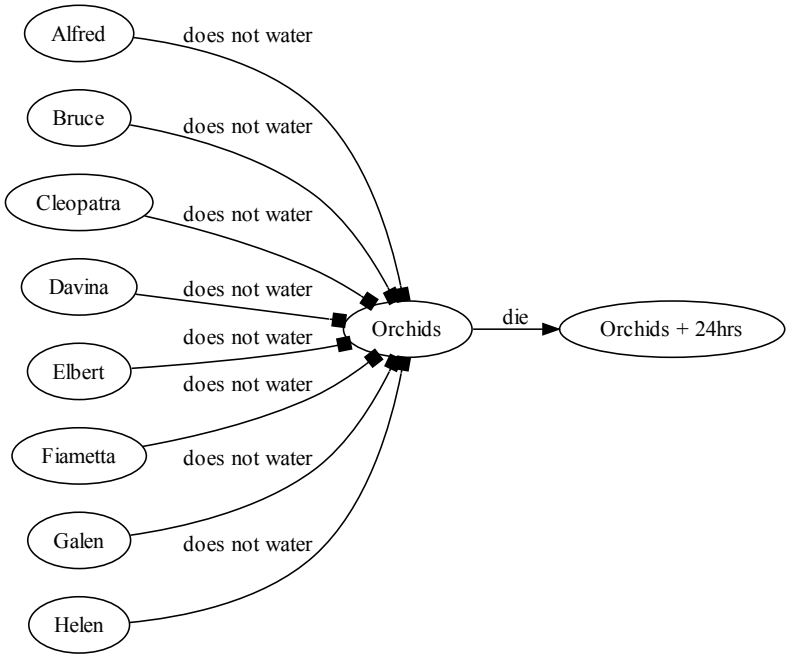

Figure 5.13: Negative causation example 3

depend upon these rather fuzzy normative conditions. This, as Beebee suggests, presents formidable metaphysical problems. I think that it raises more immediate problems too. In the spirit of the rest of this piece, I will therefore focus almost exclusively on the epistemological and methodological aspects of this issue. This means that I need either give some improved account of what these norms are that allows reliable causal assessment and inference, or I must abandon the idea of accommodating negative causal relationships.

I think the second option is impossible. If we wish to produce a theory of causality that can be useful for understanding medical causes (and I do), we have no option but to commit to negative causes. Even the most cursory study of causality in medicine suggests that negative events play an important role in causation. For example, many therapeutic strategies rely upon negative events to causally intervene. So we might attempt to reduce the incidence of HCC in a population by initiating a programme of population vaccination against HBV.[34] Similarly, negative causes play a vital role in causal inference in medicine. This was recently stated in a very clear fashion as '*sublata causa, tollitur effectus*' (Vonka, 2000, 1835), which we might rather crudely translate as 'The effect ceases when the cause is removed.'[35] For example, the prevention of HCC by vaccinating against HBV tells us something very useful about the causal structure of this case. For example, seeing any such reduction in incidence suggests that HBV

[34]I discuss interventions of this sort in more detail in 7.1.

[35]This sentiment apparently derives from the work of St. Thomas Aquinas. Yet I can find no direct instance of it in the works of the Saint. It certainly occurs in Kant at 28.573 in the Metaphysik L2: '. . . when the cause has been canceled, the effect is canceled "*sublata causa tollitur effectus*" (Kant, 1997 (1790, 336)

really does cause HCC - assuming that nothing else about the vaccination could potentially account for the effect. For instance, it might be that the vaccination campaign was carried out as part of a wider public health campaign against HCC, with surveillance measures, public engagement, actions against other risk factors and so on. Assuming that this relationship was not so confounded, the reduction seen in HCC incidence following such a specific intervention would also tell us a great deal of quantitative information about the causal relationship between HCC and HBV. For example, we might be able to estimate the proportion of HCC that was attributable to infection with HBV. From these two sorts of claims about negative causality, and the prominent role negative causes of one sort or another play in the medical literature, negative causality appears at least as important as positive causality for medical practice.

This seems to be the basis of Russo and Williamson's distinction between aetiological and therapeutic causes (Russo and Williamson, 2007, 157). Here, the process of assessing aetiological causes is concerned with those mechanisms (or components of mechanisms) that are sufficient to produce a particular state. Therapeutic causes, on the other hand, are concerned with necessary causation. Very often, therapeutics attempt to intervene on a mechanism in order to prevent the (pathological) effects by preventing a necessary, rather than sufficient, part of a mechanism.

These considerations therefore commit me to incorporating negative events as causes in my scheme. I will therefore have to try and develop this contrastive position here. While the full description of this issue will have to wait (for section 7.1), I can give a brief outline here. My account depends on three things. First, adequate functional specification of a mechanism. Second, the property of branching/track-switching that I suggest above. Third, the role of difference-making evidence. With these three things in place, scenarios of negative causation become rather more simple to understand.

Think of a correctly functionally specified mechanism for an event. As I suggested, disease mechanisms generally need to be specified in terms of a system of conceptually related outcomes. Changing a positive cause (Helen watered the orchid) to a negative (Helen did not water the orchid) should be understood as a divergence of activities at a given entity. Thus the one entity (Helen) has two possible activities (waters/does not water) as part of this specification.[36] Thus when she fails to water the orchid in a particular time-frame, the mechanism switches to the track with Helen's *does not water* activity as its causal parent. We then have to exclude the manifold non-waterings of other possible orchid waterers (Alfred, Bruce, Cleopatra, Davina...). The details of this exclusion depends upon the way that we integrate difference-making evidence with our mechanistic structure. Thus the job of the next two chapters will be to introduce such an account of difference-making, and then to show how mechanisms and difference-making can be integrated. However, briefly, I suggest that we can happily ignore the negative actions of our other possible-waterers because as individuals they make very little difference to the ultimate effect. While this means, say, that the set of possible-waterers are collectively responsible for the death of the orchid, because overall they failed to prevent it, they are individually such weak causes of the effect that we can, in general, ignore them. In other words, difference-making evidence tells us that

[36] In fact, she may well have many more than two possible activities (over-waters/waters correctly/under-waters/does not water). But I will leave this complication aside for the moment.

their (in)actions are not salient to the functional specified Helen-orchid mechanism. Absences do not mean that the productive continuity of a mechanism ceases. They instead mean a change in the routes taken through a functionally specified mechanistic structure.

5.9 A definition of mechanisms for the Russo-Williamson Thesis

By means of concluding this chapter, I would like to suggest a definition of both mechanisms and of mechanistic models suitable for the RWT.

Mechanism: a mechanism is an actively organised set of activities and entities, productive of one or more behaviours, amenable to investigation and description.

A few words of explication are in order here. First, as per MDC's account, active organisation is the key characteristic of mechanisms. Only actively organised sets of entities and activities can form a productive continuity, and it is the possession of this productive continuity that permits mechanisms to produce particular behaviours. Usually, mechanisms are polymorphous, or productive of a number of different such behaviours. In other words, the mechanism for x will often do y and z also. Mechanisms are dualistic: both activities and entities are necessary participants, and neither activities nor entities have any kind of priority over the other. Finally, as the benefits that thinking about mechanisms arises from our ability to describe them, unknowable-in-principle mechanisms - if such things exist at all - are excluded from this definition.

There are a few additional characterisations that I might suggest for medical mechanisms: these tend to be interlevel - that is, relate the operation of phenomena that are disparate in terms of scale - and display a characteristically branched structure. This comes about because medicine is often concerned with the manner of divergence of a diseased state from a healthy, physiological state.

Together, this definition of a mechanism suggests a further definition, that of an epistemic model of a mechanism:

Mechanistic model: a dependable, context-sensitive representation of a mechanism, such that the model connects both cause(s) and effect(s) of interest. This model should be capable of being integrated with other forms of evidence, in particular those that are descriptive of the overall relationship that obtains between cause(s) and effect(s) in particular circumstances. If this model displays a branched structure, it should include both structural details of these branching points, and some consideration of the antecedent events that cause the mechanism to follow one or other subsequent paths. Together, this mechanistic model should be capable of accounting for the common behaviours of the mechanism in defined circumstances that form the contextual descriptors, including those events that obtain via the prevention of the normal operation of the mechanism.

These definitions will form the basis for subsequent discussions of mechanisms and mechanistic models in the subsequent parts of this thesis.

Chapter 6

Evidence for Causality

6.1 Introduction

In this chapter, I will explore the relationship between evidence and causality. I will try to clarify a number of concepts that I have previously employed rather loosely. What sorts of evidence do we encounter in medicine? Can we legitimately make Russo and Williamson's distinction between evidence of mechanism and of difference-making? What constitutes evidence of mechanism? What is evidence of difference-making? How might we characterise them?

I will begin with McKay Illari's recent disambiguation of evidence in the RWT. This will then suggest a number of avenues of inquiry that I will explore using my viral oncogenesis examples. Here, I will show that the distinction drawn between mechanistic and difference-making evidence in the RWT is historically and theoretically tenable. I will also suggest that the sorts of evidence that are important for causation change chronologically, and that these chronological changes can be characterised in terms of the precise aspects of causal models that are being developed at the time. I will relate this to the model of discovery that I outlined in section 1.3. I will then move on, in section 6.4 to attempt a general characterisation of mechanistic evidence. While doing this, I will digress slightly in order to make a few remarks about the two unusual forms of evidence that arise from plausibility and analogy, and which played a central role in supporting mechanistic claims in the historical examples. I will then turn to discussing difference-making evidence. What constitutes this, and which strategies are effective means of generating it? What sets this form of evidence apart from evidence of mechanism? I will illustrate with some examples from the EBV and HPV cases. I then plan to introduce the reference class problem (RCP), and its relationship to evidence of difference-making. I will suggest a possible means of tackling the RCP that relies on the use of mechanistic models. This, I suggest, also presents a means by which we can seek to unify evidence of difference-making with evidence of mechanism. Briefly, I suggest that we can satisfy the reference class rule (Salmon et al., 1971) by employing features of mechanistic models. This claim will set the scene for chapter 7, where I will demonstrate the means by which evidence of mechanism and of difference-making can be integrated in support of causal claims in medicine.

A note on terminology at this point. From now on, I will adopt the term *causal models* to represent RWT-type epistemic devices which are supported by both mechanistic and difference-making evidence. These are clearly different from mechanistic models, in that they are supported by relevant evidence of

difference-making.

6.2 Two types of types of evidence

McKay Illari (McKay Illari, 2010) makes an important distinctions about mechanistic evidence.[1] The essence of her claim is that, when we think about causal evidence, we need to distinguish the type of evidence from the object of our evidence. Put another way, there is a difference between *kinds of evidence* and *kinds of things we have evidence of*. We might illustrate this by an example. We might gather evidence about the structure of a virus. But we should distinguish the evidence as type of evidence (for example, arising from X-ray crystallography on protein crystals) from that evidence as evidence of the structure of the virus. McKay Illari argues, not only that these two things are independent, but also that the RWT only supports one of them. What the RWT requires, argues McKay Illari, is evidence about kinds of things. That is, to evaluate a causal claim we need to have evidence that tells us about both mechanism and difference-making as objects of evidence. The RWT is often read as supporting a thesis about types of evidence, but McKay Illari suggests it should be interpreted as addressing the objects of evidence.

In fact, McKay Illari suggests that it may be impossible to make the mechanistic-probabilistic distinction upon evidence as it is gathered, because the two forms of evidence may arise simultaneously:

> It is in principle possible that a single experiment could establish a causal claim. Psychological work on small children investigating simple mechanical toys shows that they come to an accurate picture of the causal structure very quickly, particularly if they are allowed to manipulate the mechanism themselves... Using the framework, we can say that they acquire both evidence of difference-making (by seeing, say, that wiggling a lever moves a wheel) and evidence of mechanism (say, by observing the attachment of the lever to the wheel) almost simultaneously.
>
> (McKay Illari, 2010, 7)

She further supports this claim by arguing that our failures of classifying evidence as gathered result from the fact that doing this fails to pick out the features important to its use in causal inference. I think this is quite correct, and I will support a similar claim in section 6.4. Looking to characterise mechanistic evidence by simple methodological features does not work.[2] After all, the role that a piece of evidence plays in making a causal model of a mechanism is relative to the specification of that model. However, once this consideration is taken into account, it is possible to characterise, to an extent, mechanistic evidence by giving general descriptions of the kinds of causal facts that it is evidence *of*.[3] So I here agree

[1] She makes a number of further clarifications too, but I will concentrate on this one here

[2] McKay Illari suggests a range of possible distinctions on evidence as gathered. These include distinguishing evidence on the grounds of it being either qualitative or quantitative, relating to generic relationships or single cases, or as necessarily requiring repeated trials versus requiring simple confirmation only.

[3] I later characterise mechanistic evidence in terms of the kinds of causal facts that it tells us about. These are *activity-description*, *entity-description*, *activity-of and entity-of descriptions* and *interactivity-description*

with McKay Illari that it is only by considering the role that various bits of mechanistic evidence play in making mechanistic models that allow us to characterise them.

I will follow McKay Illari's terminology, too. She suggests that we should carefully distinguish between evidence as gathered as compared to evidence describing particular types of thing. By this distinction, mechanistic evidence (ME) and difference-making evidence (DME) are defined by the manner in which they are collected, while evidence of mechanism (EM) and evidence of difference-making (EDM) are characterised by the sorts of relationships they describe. By means of illustration, consider the following example. Say we are investigating the aetiological relationship that might obtain between infection with a virus, and the development of a particular clinical syndrome. We can distinguish the following types of evidence:

Mechanistic evidence (ME): evidence that is gathered in a way characteristic of mechanistic investigations (see, for example, Darden, 2006). In this case, we might try and form a methodological characterisation of such evidence. This might include evidence arising from electron microscopy, from animal experiments, or from studies on viral replication in cell culture. We might then seek to form generalisations about mechanistic evidence: we might say that it is generally qualitative, that it generally does not require repetition in order to be significant, that it generally concentrates on the operation of micro-level processes, and that it generally arises from laboratory investigation, for example.

Evidence of mechanism (EM): evidence that tells us about the particular mechanism by which the virus mechanistically acts to cause the disease. For example, we might find that the virus attacks certain types of cell, thereby causing them to rupture, in a way that explains particular features of the clinical syndrome (say, the presence in the blood of particular cellular components). We might then be able to characterise this evidence in a contrastive way from that characterisation of mechanistic evidence: evidence of mechanism tends to address the manner in which particular causal processes are organised, for example. I will return to this characterisation later in this chapter.

Difference-making evidence (DME): as with ME, we might characterise difference-making evidence in terms of the way in which it is gathered. Examples of difference-making evidence in this case might include observational trials that attempt to find a correlation between infection with our virus of interest and the development of the relevant disease, or an interventional trial that attempts to discover what happens to the clinical syndrome when an effective vaccination against the virus is given to a population. This might lead us to suggest general qualities of such evidence: that it is often quantitative, that it concentrates on the relationship that obtains between the cause and the effect, or that it generally arises from population-level investigations on the phenomenon of interest.

Evidence of difference-making (EDM): finally, evidence of difference-making tells us about the particular difference that our cause makes to our effect. Here, we might find that infection with the virus accounts for only a proportion of our clinical syndrome (suggesting the operation of other difference-makers), or that our clinical syndrome is just one of several different individual-level effects of infection with our virus (suggesting the operation of other pathogenic mechanisms resulting from infection with the virus for future investigation).

6.3 Evidence in the cases

I propose to give a rather illustrative, chronological account of causal evidence in this section using the EBV case. How does evidence for causality arise, and how does this evidence alter our causal model? I will use the distinct periods of causal discovery that I first applied to the case studies to distinguish the changing types and uses of evidence as they come about. To recap from section 1.3: while discussing the EBV and HPV case histories, I suggested that four stages of causal inference were apparent. First, each story began with the **suspicion** that some sort of stable, causal system was responsible for producing the disease in question. Next, the **causal domain** of this system was sought. Understanding the structure of this system then comprised the next step, **mechanism construction**. Finally, this new causal model was **applied** in a way that attempted to cast cognitive and therapeutic light on the disease. In the following section, I will attempt to give an evidential characterisation of these four periods. With the exception of the very earliest stage, I will also relate these general features of causal discovery to the RWT.

6.3.1 Evidence and the suspicion of causes

I suggested earlier that this phase of causal discovery is essentially characterised by the suspicion that a particular clinical phenomena might arise as the result of some kind of discoverable causal mechanism. In large part, this entails showing that a particular set of clinical findings arises because of the operation of some sort of stable disease mechanism.[4] In the EBV case, this phase essentially consists of Burkitt's initial clinical description of the lymphoma syndrome, and the rather speculative claim that these clinical findings came about as a result of some kind of (potentially interpretable) causal mechanism. This phase is, essentially, a process of syndrome recognition. A small number of clinical features - symptoms, signs and investigation results[5] - were noted to occur in regular association, and it was this regularity of co-association that suggested the operation of a stable causal process. This is the rudimentary outline of what it means to suspect a cause. As such, I will not give an RWT-compatible description of what it means to suspect a cause.

I will give a more detailed account of this process in the EBV case below. While I suspect a similar thing happened in the cervical cancer example, the details are rather more opaque, occurring as it did in the poorly documented prehistory of medicine. However, elements of this suspicion are to be found in the early accounts that differentiated cervical from uterine body tumours (section 3.1.1), and in those early publications that attempted to detail the behavioural peculiarities of those social groups at high risk of developing the disease (section 3.2ff). Given these historical difficulties, I will focus on the EBV case for the sake of clarity.

What is the first part of the story? Burkitt first observed one child with an unusual jaw tumour which could not be easily explained by a known clinical syndrome. However, he did not yet appear to regard this single case as an anomaly in need of explanation. He instead - initially - treated it as a simple curiosity. It was not until the discovery of a second, very similar case, that these observational

[4] This point is rather obvious, but it can be a significant part of understanding disease causation. For example, in the McArdle case, the vital finding that this set of apparent symptoms were real, rather than spurious, and that they occurred as the result of a disease state took a great deal of investigative effort.

[5] For reasons of clarity, I will generally refer to these sorts of clinical findings as *symptoms*

findings began to suggest the possibility of a causal explanation. There is some historical support for this claim. For example, Burkitt very clearly described the psychological differences from seeing the first two cases. While the first was 'baffling' (Burkitt, 1983, 1777), it was not until he saw the second case that his '...interest was rivetted immediately' (Burkitt, 1983, 1778). This suggests to me that it was not simply the novel and unusual features of either of these two cases that made them interesting. It therefore seems incorrect to claim that the psychological impact of either of these anomalous individuals alone was responsible for suggesting the operation of an interesting causal process. Yet I also do not think that this case represents a good demonstration of the psychological power of Humean constant conjunction. Finding a group of odd symptoms together more than once plausibly suggests the operation of a common causal process.[6] But in this case I suggest that it was not simply some kind of psychologically impressive induction across these cases that led to the search for a (causal) explanation either. It was, instead, the novelty of the purported *syndrome*, rather than that of the single cases, that did the work of surprising Burkitt.

I can provide some further support for this claim. Induction over cases alone appears historically insufficient to explain the way that events came about. Both Burkitt and others had previously noted the high incidence of unusual jaw tumours at Mulago without previously coming to the conclusion that they constituted instances of a particular syndrome in need of causal attribution (see section 2.2.3). Similarly, many other authors reported individual cases of these odd jaw tumours that they did not subsequently go on to investigate further. Now this seems troubling, and in need of some explanation. I propose to account for this difference between Burkitt and his contemporaries by drawing a (causally important) distinction between a clinical syndrome as opposed to a group of cases.

This term clearly needs some elucidation. I suggest that a syndrome is best interpreted as a distinctive, evidentially complex clinical entity that occurs in a phenomenologically regular manner. This is a definition that does not necessarily depend upon knowledge *of* or *about* an underlying causal process. Down syndrome, for example, was known and phenomenologically well understood for some 90 years before its causation was understood at all (Down, 1866; Lejeune et al., 1959; Einfeld and Brown, 2010). However, the regularity with which the features of a syndrome co-occur does serve as evidence of the possibility of some common cause. But I do not think that this regularity can be used to ground the syndrome-cases distinction. Instead, the difference between sets of symptoms and syndromes is a methodological one which arises from the differing way in which they are clinically recognised.

Syndromes are not simply aggregates of symptoms or other clinical findings. Indeed, trying to firmly construct definitions of syndromes by groups of symptoms is likely to be problematic. Part of this difficulty comes from the subtlety of features of the syndrome that may play an important role in the diagnostic process, as I go on to explain. Further aspects of this difficulty might also arise from the lack of clear necessary or sufficient relationships that symptoms and syndromes share. I suggest that it is usually the case that no particular symptom is a necessary component of a syndrome. Likewise, there is generally no symptom that is sufficient evidence for a syndrome.

[6] Generalised Reichenbach common-cause principle

Yet here is the paradox: syndromes are distinctive in a way that sets of symptoms are not, and it is a relatively unproblematic business to detect a syndrome in a particular individual, even though our understanding of what - symptomatically - constitutes a syndrome is so weak. The Diagnostic and Statistical Manual (DSM) provides a good operational example of this distinction.[7] A study of this diagnostic manual for psychiatry reveals a relatively small number of atomic symptoms or clinical findings that can be used in the diagnosis of psychiatric disease. More interestingly, most combinations of symptoms do not fall cleanly into one particular diagnostic category. Thus the business of making a diagnosis by making a list of symptoms and then comparing the list of symptoms to the DSM categories is likely to bring up a list of a great many possible diagnoses. Yet typically the diagnosis of, say, depression is not problematic in this respect in clinical practice. In fact, confirming a suspected diagnosis by referring to the features in the DSM may show a rather good fit between observed and required symptoms. How can this be the case?

I suggest that the best candidate for explaining this apparent difficulty is to consider more subtle features of clinical syndromes. While, say, the manner in which an individual avoids or prolongs eye-contact may be quite plain to the observer, recording this manner may yet prove rather challenging. That is, syndromes may have evidentially important phenomenological characteristics that defy analysis. It may be that these features are temporally inconvenient (say, very short duration facial movements), that they are apparent in a sensory modality which lacks a good descriptive vocabulary (sound and smell, for example),[8] or that they are only suggestive of the syndrome when seen in collaboration with other, similarly subtle, features and so on. Rather as you might find it hard to describe which features allow you to recognise an old friend at a distance, in low light or over a poor quality telephone line - yet it is easy to know who it is - the actual evidence used to recognise a syndrome may well differ dramatically from our representation of that evidence.

So I do not mean that these bits of evidence that support the recognition of a syndrome in an individual are in any way ineffable or unknowable. It is just that these features may well be unlikely to survive the translation to their conventional written, pictorial or numerical representations. In other words (perhaps like the old friend) the instances of a syndrome ‘smell’ the same, and are not simply examples of similar clinical findings. They are instead something that is indicated by the recognition of obscure, poorly-characterised symptoms. Thus, when only the clear and present symptoms are recorded, the discovery or diagnosis of a syndrome appears to assume the form of a gestalt recognition.[9]

[7]The DSM (American Psychiatric Association, 1994) lists a great many psychiatric conditions, along with a list of clinical features. Diagnosis - according to the manual - can be achieved by meeting a certain number of diagnostic features of the disease. Its construction and use in clinical practice are currently rather controversial.

[8]The terminological confusions that surround the written and pictorial representations of lung sounds are an excellent example of this fragility. It may be the case that there are other reasons that important features of syndromes are difficult to represent. For instance, certain conditions are associated with embarrassing, yet diagnostically distinctive, symptoms. It may be the case that clinicians are reluctant to investigate or describe these symptoms for fear of upsetting patients. For instance, sufferers of isovaleric acidaemia have a characteristic smell, which is said to resemble sweaty feet, or old socks (Tokatli et al., 1998).

[9]I use the term to indicate that the recognition that something might be a syndrome, or that an individual might have a particular syndrome, appears to occur as a whole. As I have suggested, though, I think we can consider more specifically the pieces of evidence that contribute to this process. However, the representational constraints in play may complicate the business of speaking

This tension between the evidence used, and the representation of the evidence used, has a possible origin in the professional conventions of medicine. Representations of evidence, learned as part of medical training, constrain the sorts of evidence that are discussed when it comes to discussing diagnoses. But they do not necessarily constrain the evidence considered when the clinician is actually making a diagnosis. In fact, there is a tension here in what is learned at different parts of medical training. While the formal representations of clinical encounters serve to negatively constrain the types and manner of evidence used, the way that clinical training actually occurs seems to facilitate just such tacit evidential considerations. The many years spent, under supervision, interacting with patients, attempting to make diagnoses and comparing physiology with pathology in manifold ways serves to familiarise the clinician with these gestalt features of disease in a way that learning lists of disease features would not. It is familiarity with patterns of disease that grant a clinician diagnostic skill. In turn, I think that it is highly likely that this profound familiarity with the patterns of disease not only leads to diagnostic competence, but also grounds the possible recognition of syndromes.

Recognising something to be an novel anomaly in need of explanation necessarily requires syndrome recognition. Simply meeting an individual with an odd set of symptoms remains utterly unremarkable,[10] and is unlikely to provoke the sort of response that happened in the BL case. Put another way, symptoms do not tend to provoke research, while syndromes do. As with Kuhn's description of anomaly recognition (section 1.3), this recognition of a novel syndrome as a prompt to research is not an inductive process. For example, McArdle (McArdle, 1951) was prompted to carry out extensive investigations on the basis of a single case. Why it took Burkitt twice as many cases seems unimportant. In any case, carrying out a useful induction from two samples seems almost as risky as that from the single case.

In conclusion, how might we rationally reconstruct this early part of the EBV story. First, Burkitt encountered two patients with distinctive features. This led him to a complex process of syndrome recognition, whereby the clinical features of the two cases contrasted with his education, background knowledge and experience (sections 2.2.3 and 2.2.4), leading him to conclude that these individuals represented instances of a (possibly novel) syndrome. As syndromes are peculiarly psychologically striking, finding something that resembled a novel syndrome in an individual was more compelling than finding a new set of symptoms. My assertion is that this makes the process of syndrome recognition in these two children the driver of causal suspicion. What it does not do is to say very much about the nature of the cause(s) operating. Thus, both parts of my account of causal suspicion critically depend upon Burkitt's education, background knowledge and clinical experience. The tacit nature of this knowledge

specifically about what precisely these features are. I therefore think it is wiser to treat this phenomena as occurring as a whole, at once.

[10]This is a feature of medicine that is rather overlooked. It remains the case that a great many clinical interactions remain diagnostically and therapeutically unsatisfactory from any perspective. It is pretty uncommon to meet an individual (un)fortunate enough to have developed a 'text-book' case of any disease. Common diseases often manifest in unusual ways, many symptoms remain resistant to useful explanation or intervention and there are so many uncommon, unexplained medical problems that much of what is encountered in practice cannot be either unified or causally explained. This was especially likely to be the case for Burkitt, given the geographical separation between his training and his practice.

and experience is therefore a possible source of the historical difficulties of elucidating the precise manner in which this discovery came about. However, we can point to some differences between Burkitt and those others who encountered jaw tumours but did not attempt to investigate them by drawing this syndrome-cases distinction. The latter treated the individuals with these lesions as instances of jaw tumours *simpliciter*. Burkitt, on the other hand, suspected that they were suffering from the same clinical syndrome. So these jaw tumours were, for Burkitt, striking because they were potentially manifestations of an unknown syndrome.

I do not want to argue that the RWT can account for this first stage of causal discovery. However, once the operation of a cause is suspected, the RWT certainly can be used to interpret the detailed development of a new cause. So I will relate the following stages of discovery to the RWT.

6.3.2 Evidence and causal domain-finding

This phase of EBV-BL research began with a regular, syndromic anomaly in need of causal explanation. The next step was research that looked to demonstrate the approximate way that plausible causes of this syndrome might operate. I suggest that this means that researchers attempted to determine the domain of the suspected cause. At the end of this phase, therefore, researchers understood the likely, rather than actual, nature of the causes at work producing the suspected syndrome.

Because our models of causes are contextual (see section 5.4.4ff), this stage of causal inference is therefore largely concerned with what a mechanistic theory of causality might call functional specification. What does the cause do? Over what sorts of entities and activities is the mechanism likely to operate? In the EBV case, I suggest that at the end of this stage of discovery, researchers knew that BL was a malignant disease that caused a characteristic clinical syndrome. They also suspected that it was caused, at least in part, by the actions of an infectious agent - probably a virus. The types of evidence employed to reach these conclusions are, thankfully, rather more amenable to retrospective analysis than those at the earliest stage of discovery. Of particular significance are clinical descriptions of the syndrome, histological examination of tumour material, comparisons of this syndrome to those produced by other, known disease mechanisms, and simple epidemiological data. In the cervical cancer case a similar process of syndrome description was also evident. Risk factors for cervical cancer were elucidated; the similarity of these risk factors to those of an infectious disease were noted, and possible causal agents were suggested that appeared capable of accounting for these risk factors.

Thus, in general terms, this stage of discovery involved a process of exploring two aspects of the purported causal relationship. First, researchers looked to develop the extreme right-hand edge of the mechanistic model[11] with its descriptions of final states of affairs. This was constituted by developing a detailed phenomenological, conventional and canonical description of the clinical syndrome associated with the disease state.[12] This sort of description included symptoms, signs, investigation results (histopathological, blood-tests, roentgenography) and basic epidemiology. As an important component of this description, researchers began seeking plausible ways of reconciling these features. This pro-

[11] This is a terminological difficulty. Like MDC, I am resistant to thinking about mechanistic inputs and outputs.

[12] c.f. the manner in which McArdle's description of McArdle disease rapidly became the 'classic' clinical picture of that disease.

cess of unification was both internal, between different clinical features, and external, between clinical features and their causes. As an example of internal unification, an histological finding might plausibly explain a symptom or a haematological finding. As an example of external unification, an histological finding might be plausibly explained by suspected pathological features of the disease state.

This broad, plausible description of the causal framework for the disease state forms that which is to be explained by the provision of a causal model. This entails finding, like the Amazon, the nature of the 'source' of plausible causal mechanisms. I interpret this, on the grounds of the historical cases, as the search for the significant points of branching from the normal physiological mechanism that could, plausibly, explain either syndromic aspects of the disease state, or the causal antecedents of these clinical descriptors.[13] Importantly, this stage of enquiry also involves the comparison of the morphology and features of this disease state with others known, to make sure that this is not, say, an odd manifestation of a known disease.

This proto-mechanism undergoes significant development as the domain of the cause is sought. First, further syndromic features are elucidated. This includes finding additional symptoms, signs and abnormal investigation results associated with the syndrome. A degree of causal-mechanistic attribution may also take place. For example, in the BL case, the common histopathological features of tumours at different anatomical sites, or of tumours in different individuals, suggested that a similar process of oncogenesis was occurring in each distinct instance. Thus, a range of syndromic features were causally unified into instances of the same oncogenic process.[14]

Next, a number of plausible alternative causal disease models were examined. One very clear example (in the BL case) was the comparison made between the tumour syndrome and leukaemia. But a mechanism that could account for the clinical syndrome that developed around BL would have to be an extremely odd one. As Burkitt noted (Burkitt, 1983, 1777), the clinical features of these patients were aetiologically 'baffling'. No known pathological process seemed to provide an easy, plausible explanation. He further noted that distribution across all four jaw quadrants seemed to exclude the usual explanations for bony swelling: neoplasia and infection.[15] So no existing disease state seemed to

[13] See section 5.6. I will later suggest that it is branching because of setting effects that is the cardinal feature that is explained.

[14] I will not say very much about this. I think this business of local causal attribution proceeds in the same way as the global one. However, each individual bit of causal inference, owing to its causal simplicity and plausibility, is both much easier to perform, and less psychologically impressive than the eventual, overall one. This lack of drama, and chronological brief-ness, therefore renders the internal mechanics of the manner of causal inference for these simple mechanistic links much less clear to the historical investigator than those of the main mechanism. But this historical difficulty does not challenge my thesis that causal inference happens in a recursive way

[15] As a matter of interest, I have looked up the differential diagnosis for bony swellings in the jaw (Bouchier et al., 1996, 335-9). The differential includes infections (either of dental structures or of bone), trauma, dental and gingival pathology, genetic syndromes (familial fibrous dysplasia), salivary gland swellings and a range of tumours. Benign tumours include jaw osteomas, ossifying or cemenifying fibromas, giant-cell granulomas, ameloblastomas and myxomas. Malignant tumours include both primary bone tumours and metastatic disease. Because most of these diseases are well localised and circumscribed, few of them could have offered a potential explanation for the clinical features observed by Burkitt. There are a couple of exceptions. First, while osteomas of the jaw are generally singletons, multiple osteomas may occur as part of Gardner's syndrome. This, however, is very rare and the main effect of the syndrome is to produce multiple colonic polyps with malignant potential. It was described before Burkitt's work (Gardner, 1951). A confusion between it and the clinical features of BL seems rather unlikely. Second, fibrous dysplasia of

present itself as a good fit with the syndrome. When considered along with the clinical features (and their interrelations), this provided good evidence that a new causal mechanism, capable of explaining a process of simultaneous multifocal oncogenesis, occurring in a geographically constrained manner, in an unusually youthful group of patients, would be required.

In RWT terms, this represents evidence of difference-making. Something made a difference to this group of children in East Africa, in contrast to those in other parts of the world,[16] with other types of disease (leukaemia), in patients of other ages (age-incidence curves) or with normal human physiology. The contrast class, against which these findings demonstrated the existence of a difference, also act to constrain the types of possible mechanistic models that might account for it. For instance, tumour homology between sites and cases implied the likely existence of a common cause that was of a broadly oncogenic type. Likewise the geographical restriction of the tumour indicated that this part of the mechanism was, perhaps, the result of a geographically determined factor, such as an environmental toxin, a behaviour or a pathogen.

In summary, then, evidence of both mechanism and of difference-making were in play when the causal domain for BL was sought. The early pieces of difference-making evidence were the characteristic geographical findings and the age-incidence relationships. The mechanistic findings were the contrastive examinations of the syndromic features of BL with those arising in other disease states (i.e. mechanistically demonstrating that the tumour syndrome was not the result, or not a likely result, of a known mechanism), the histological details (offering both positive and negative constraints upon the mechanism), further details of the clinical syndrome, and their manner of interrelation, and the possible mechanisms underlying the geographical and age-incidence findings. Thus a range of simple observational evidence, together with a complex web of background knowledge, supported the beginnings of a causal mechanism for BL.

In the case of cervical cancer a similar process occurred, although it was tempered by (phenomenologically good) knowledge of the clinical syndrome of cervical cancer that was already in place at the time that it occured. Suspicion, in this case, consisted of finding the sexual risk factors for the disease. Domain finding, on the other hand, was a process of determining the types of entities that causally underwrote these risk factors, relating these in a plausible fashion to the clinical picture, and determining the domain of the causal structure linking the two. Essentially, therefore, this debate was focused on the clear discrimination of cervical from uterine corpus tumours, the determination of which risk factors appeared to make a difference at a population level, and the analysis of these risk factors in terms of their plausible components. Moreover, the elucidation of these risk factors suggested (fairly early on) that the correct aetiological domain involved infection with herpes viruses. Domain finding therefore involved, in this case, extensive analogical reasoning with other types of malignant disease caused by herpesvirus

---

the jaw could have accounted for some of the symptoms seen. This is a rare disease affecting children in which bone is gradually replaced by fibrous tissue. This, while it could potentialy affect all jaw quadrants simultaneously, tends to manifest as a non-focal enlargement rather than a tumour. The low incidence of the disease also counts against it as a good explanation for Burkitt's first two cases.

[16] Burkitt was already aware that tumours of the jaw were considerably more common in Mulago that in Europe or America (Singh, 1955, 70). We also have various reports of unusual jaw tumours recorded before Burkitt arrived in Uganda (p. 31ff).

infection. I discuss the role of analogy as evidence for causation in section 6.4.1. There I will defend the assertion that this sort of reasoning leads to a mechanistic type of evidence. As I have already suggested, though, other types of mechanistic evidence were lacking in this case. In this period, therefore, the majority of domain finding research in the HSV case involved the analysis of difference-making evidence.

6.3.3 Evidence and the making of mechanisms

In both viral oncogenesis cases, this phase of inquiry had a characteristic evidential appearance. Rather than investigating the possibility that the tumour could be caused by a virus, researchers instead attempted to discover which particular virus was causing the tumour, and the underlying mechanism of this causal relationship. An analysis of this stage will therefore concentrate on the manner in which appropriate models of causal structures, capable of providing suitable cognitive leverage across a range of specific situations, were developed. The task of this section is therefore to show what part evidence played in this process. What I will not do here is to reprise the large number of well-developed accounts of mechanism construction that are already available (see, for example, Darden, 2006; Craver, 2007; Bechtel, 2008) What I will do instead is to use the case histories to illustrate the distinction between kinds of evidence along McKay Illari's lines, as introduced in section 6.2. That is, I will show that there is a distinction between kinds of evidence and the kinds of things we have evidence of. With this distinctions in mind, I will use the EBV case to give a broad outline of this part of my account.

At this stage of the BL research programme, the focus of the investigation changed. From attempting to discover both the phenomenal nature of BL and the *types* of entities responsible, researchers began to work towards giving an account, in specifically causal terms, of the precise nature of the causal structure connecting the two. Thus the methodologies and epistemologies at work were all based, fairly explicitly, on the earlier plausible assumption that the tumour was caused by a virus (see section 2.4). This working assumption was methodologically essential. It granted researchers legitimate access to the techniques, equipment and patterns of research from virology, oncology and non-human tumour virology that would otherwise have been unavailable. Epistemologically, it also offered researchers the opportunity to gain cognitive advantage by employing general causal mechanisms and structures, shorthands, argots and technical talk from each of these fields. As a consequence of the novelty of the BL research programme, these first steps, with a grounding in each existent research programme, formed the paradigmatic basis for human tumour virology as an independent, interdisciplinary research programme.

Chronologically, the very first steps in this case were purely virological. A detailed investigation, employing a range of sophisticated virological techniques, were conducted using tumour biopsy material (see section 2.4.1.2ff) However, these attempts initially failed. No virus was found despite extensive, methodologically pluralist investigations of tumour material. While this failure provided little direct advantage to BL research, it has since become an fundamental tenet of tumour virology that tumour viruses may be extremely difficult to find *in vivo*. This had significant repercussions for the investigations surrounding both cervical cancer and HCC.

It was the accidental discovery of the culturability of BL cells, that provided the first (positive) steps

in the BL programme. While this ability to grow in culture was an unusual property for lymphoid-line cells, it was fairly well-described in other tumours (see section 2.4.2.1). The subsequent discovery of the general culturability of BL cells, and the production of multiple BL cell lines, suggested that - no matter what the nature of the causal mechanism giving rise to this behaviour - BL cells possessed this ability *robustly*. No matter who the tumour arose in, or where it arose in them, the tumour cells had the ability to grow in free culture. This piece of evidence of difference-making thus arose from a contrastive comparison of BL cells in culture versus other tumour cells in culture. It required detailed knowledge about how cultured malignant cells usually or typically behave. Thus this finding required knowledge of a range of oncological techniques and general causal models in order to come about.

But this (oncological) piece of evidence of difference-making was rapidly joined by a (virological) mechanistic one: typical tumour cells are not usually packed with viral particles at many different stages of maturity (section 2.4.2.2). This was evidence of mechanism regarding the manner in which the virus caused the tumour. Again, it only became so when interpreted from the virological stance: the virus was the oncogenic driver, rather than a passenger, by analogy with the properties of other viruses, and their manner of disease causation. Thus EBV became part of a causal mechanism capable of accounting for the disease state.

This causal-mechanistic role gained further support from cell transformation experiments. EBV, it seemed, could be the relevant difference-maker in malignant cell transformation *in vitro*. This difference-making result, when folded into the developing EBV causal model, becomes an important piece of evidence of mechanism. The difference that EBV exerted on cultured cells, as a whole, became just part of the more complex causal structure that explains the development of BL in real, complex, populations.[17]

In turn, this cellular transformation became mechanistically decomposed in terms of the activities of viral entities upon cellular mechanisms. For instance, the typical chromosomal abnormalities produced in these cells were described, and became entitified; the lack of lytic effect - unusual for a herpes virus - produced by this agent in the test systems (Epstein et al., 1965a, 70-1); the process of integration of the viral genome into that of the host; and so on. This is the nature of recursion seen in causal inference: differences become mechanistically explained, mechanisms become functionally related to difference-makers. Meanwhile, other causal-mechanistic candidates were excluded by a related process. While possible, plausible causal mechanisms implicating these candidates in BL oncogenesis were available, the corresponding actual difference-makers, or mechanistic findings, were not.

In a similar way, the ubiquity problem (section 4.4.1) was discovered, and resolved. The role of EBV in other diseases arose from serological investigation. This was methodologically and epistemically a virological approach to understanding disease.[18] However, the range of control sera chosen shows an oncological colouring of this experimental instance. For those conditions in which EBV had not previously been implicated this was difference-making evidence. EBV seroconversion appeared to make a

[17] Cell transformation experiments are distinctively oncovirological, and neither oncology nor virology alone can lay claim to this experimental practice

[18] Although since that time the role of immunology in understanding malignant disease has become more commonplace. Perhaps this is a further example of the wider impact that the understanding of the EBV-BL story has had on medical research.

difference to the development of these diseases. The resolution of this problem, though, depended upon the mechanistic distinction of the manner in which the virus was causally implicated. In other words, EBV was a difference-maker for all these conditions. But the mechanistic aspects of the developing causal model meant that researchers could, in detail, tell a story about the differences between the differences. Without mechanistic evidence and the construction of a range of mechanistic models, the problem would have remained insoluble. And so on.

In summary, techniques lifted from a number of medical disciplines were employed in the construction of a distinctly oncovirological causal model implicating EBV in the development of BL. Electron microscopy of biopsy material was virological, not oncological. But cell culturability was oncological, not virological. Immunological techniques were (until very recently) distinctively virological; histopathological examination of cells oncological.[19]

Depending on the context, each piece of causal evidence can be interpreted in different ways. For instance, a given piece of evidence can be, plausibly, interpreted as either mechanistic, or difference-making evidence. The precise way in which a given piece of evidence is interpreted depends upon subtle contextual factors that influence the functional specification of the piece of the causal structure under examination. For a given type of specification, those types of evidence that relate a causal feature to the (level-relative) properties of the part of the causal structure under examination serve as mechanistic evidence. Those that relate behaviours to one another in a stable manner are difference-making. But these roles are, necessarily, relative to the precise bit of causal structure under test at that time. So evidence that originally appears to be that of difference-making (say, the serological correlation between EBV and non-BL disease important in the formulation of the ubiquity problem) within one context (in this instance, that of a programme to determine the causal specificity of EBV for BL) can later play a mechanistic role in an alternative mechanistic model. In this specific example, this occurred in the construction of an overall causal model for the roles of EBV as a human pathogen.

6.3.4 Applying evidence

The causal structure relating EBV and BL was still incomplete by the early 1970s. While the mechanistic model component of the structure seemed highly plausible, there were a number of significant problems with the view that this model was actually happening in the BL population. Arguments over a range of issues (reviewed in section 4.4.1) led to much resistance to the causal model as a whole. This phase of inquiry therefore consisted of resolving these problems by demonstrating that the EBV-BL mechanistic model did actually occur in BL populations. This part of causal discovery was, therefore, concerned with evidence of difference-making, and the manner in which it related to mechanistic models. I review this process in detail, along with the similar process that revealed that HSV was not a cause of cervical cancer, in section 6.5.

[19] As an aside, perhaps this lifting and modification of techniques is a source of local incommensurability in models of causal structure. Each bit of causal structure intimately depends upon the technology, methodology and epistemology of those constructing it. Important differences exist in these areas between disciplines.

6.4 Evidence of mechanism

What constitutes evidence of mechanism? That is, what sorts of experimental or observational findings are likely to participate in the formulation of a good mechanistic model? Is it possible to characterise either mechanistic evidence (ME) or evidence of mechanism (EM)? Can we do this methodologically? Epistemically? Ontologically? How do these sorts of evidence depend upon the context under which they are sought? What is the nature of ME and EM in the viral oncogenesis case-studies?

Mechanisms and their models are concerned with causal dependencies. A model of a mechanism should illustrate just how a regular arrangements of causal antecedents causes a particular entity to act in a particular way. EM, therefore, has to give us some insight into how these causes connect together. Equally importantly, EM must tell us how - under certain conditions - the interactions of these entities and activities gives rise to the properties of the mechanism - productive continuity, stability, robustness, polymorphism and so on. Because of these various demands, and those distinctions between object and type of evidence outlined by McKay Illari (McKay Illari, 2010), simple methodological definitions of ME appear unlikely to be adequate means of providing a useful characterisation of EM. First, it is simply not the case that EM is just causal evidence that arises from *in vitro* investigation. Nor, I think is it possible to characterise EM as necessarily arising from laboratory investigations. I think that arising in a laboratory is neither necessary nor sufficient to make a piece of evidence distinctively mechanistic, in either sense. Similarly, but in Russo and Williamson-like terms, neither EM nor ME can be characterised as being simply non-probabilistic causal evidence. Again, some evidence of difference-making is non-probabilistic (as in the McArdle case), while some important evidence about mechanisms found in the viral oncogenesis case studies is absolutely and innately probabilistic.

There are, instead, a number of characteristic features of mechanistic models that seem to arise from certain types of evidence as gathered. This suggests that it might be possible to characterise EM in terms of the features of the ontic mechanism it purportedly describes. First, some mechanistic evidence looks to describe the nature of entities and activities such that they can be included into a mechanistic model. For instance, researchers looked to demonstrate the morphology of EBV, described the condition of host-cell genomes in EBV-transformed cells, described the structure of muscle glycogen in McArdle patients and so on. All of these were instances of descriptions of (possible) causal entities. There were similar descriptions of causal activities: the means by which EBV infected cells, the type of exercise under which the characteristic symptoms of McArdle disease manifested, and the role of various early HPV proteins in disrupting cell-cycle regulation. But this is a fairly weak definition. It seems far too strong to suggest that any research that deals with parts of a causal mechanism in detail is distinctively mechanistic. However, there are more telling characterisations of EM. Distinctively, EM gives descriptions of the nature of the *relationship between entities and activities*. A range of experimental practices give rise to evidence that falls into this class.

First, some experiments attempt to determine the ways in which a given entity acts. Thus researchers sought to determine the range of conditions under which EBV would enter either the dormant or lytic phases of its life-cycle (section 2.1.3). Similarly, HPV researchers sought to determine the host-

cell tropism of different types of HPV on the grounds of their genetic sequence and other properties. The relevant conditions under which such an entity would act in a given way were sought. I term this sort of strategy *activity-description*. Second, researchers looked for the entity, or entities, that were the producers of certain activities. A good example of this is the detailed investigations that subsequently took place to detect the precise means by which the range of early HPV proteins caused cell-cycle deregulation. This activity was found to arise by the (different) mechanistic actions of E5, E6 and E7 upon cellular machinery (outlined in section 3.6.3). I term this sort of strategy *entity-description*. Third, in the case where entities were suspected to act on an unknown target entity, investigations were performed to try and find out what lay downstream of them. Again, the elucidation of the specific functions of HPV E5, E6 and E7 are good examples of this. While the cellular-level activity of the expression of these proteins (immortalisation) was well understood when they were first causally implicated, the actual molecular-level targets of these proteins were not known. Therefore research consisted of finding out which entities were the targets of the activities of these entities. I term this type of research *activity-of and entity-of description*. Fourth, the heuristic of seeking intra-mechanistic interactions - in the sense of sections of the mechanism that had synergistic or antagonistic effects on other parts of the mechanism - was applied. We might think of this as *interactivity-description*. The production of evidence that allows the telling of such a detailed causal story seems distinctively mechanistic. So this is not simply evidence regarding the *existence* of entities and activities, nor is it evidence revealing of the ways that these participate in simple causal sequences, but instead deals with entities and activities as embedded in complex causal models. For example, how do the range of viral proteins act - in concert - to effect cellular transformation. How does the picture change if one - or some - of them is altered? Inhibited? Increased? Does an entity really act stochastically, or is it instead the case that it acts deterministically under certain complex arrangements of other causal factors? Evidence that tells us about the real complexity of the workings of biological mechanisms in a way that allows us to construct mechanistic models is distinctively mechanistic.[20]

This suggests a third means of classifying something as evidence of mechanism. While both mechanisms and mechanistic models are distinctively interlevel, and medical mechanisms are characteristically extremely complicated, EM tends to speak to the manner in which very short causal links come about. These tend to operate at one particular, or between very closely related, levels of organisation. This seems paradoxical, until it is remembered that the job of EM is to describe the internal workings of the mechanism by means of mechanistic models. What it does not do is to describe the *overall* behaviour of the mechanism.

While I have suggested that very simple methodological characterisations of ME are not likely to be operationally useful, there are some more nuanced methodological features of ME that seem likely to prove more fruitful. The generation of ME generally takes place within fairly well defined disciplinary borders. While the entire set of mechanistic evidence acting to support, say, causal claims about the relationship between EBV and the clinical development of BL appears methodologically utterly disparate,

[20]This sort of causal complexity appears to result in the manifold representational complexities that appear in medical contexts.

when it is divided by bioscientific discipline a more congruent picture emerges. For example, the use of electron microscopic techniques employed to understand EBV morphology all relied on similar tools, techniques, epistemologies and background knowledge as the non-specific virological procedures for understanding the morphologies of viruses in general. This kind of disciplinary reference plays a very important role in determining the most fundamental level at which ME is sought, and hence to the level at which the resulting mechanistic model bottoms-out (see section 5.4.3).

I will now digress slightly to discuss a critically important type of mechanistic evidence that played a major role in all the historical cases referred to. This is the evidential role of plausibility and analogy in causal inference and understanding.

6.4.1 Plausibility, analogy and mechanisms

Plausibility and analogy played a vital evidential role in causal inference in both of these cases of viral oncogenesis.[21] These causal structures were seen as highly implausible at the time of their construction. Yet there is an interesting tension here: mechanisms are very often described as merely plausible descriptions of the way that things happen. This role for mechanisms is explicitly present in many medical schemes of causation.[22] For instance, Surgeon General (1964, 182; 185ff) features a requirement for determining the 'Coherence of the association' when determining causation. Likewise, Hill (1965, 298) cautions us that '... the cause-and effect interpretation of our data should not seriously conflict with the generally known facts of the natural history and biology of the disease.' Evans (1976, 192) suggests that 'The whole thing should make biologic and epidemiologic sense.' Finally, Elwood (1988) demands 'Are the results plausible, in terms of a biological mechanism?'

I think these suggestions, that mechanisms are simply plausible descriptions of how things happen, are mistaken. They seem to be operationally problematic too. In fact, a nuanced reading of the relationship between plausibility and mechanisms is rather complex. By way of illustrating the relationship between those mechanistic models that are primarily constructed using plausibility and analogy (*plausible mechanisms*) and mechanistic models that have some more concrete and direct evidential support (*empirical mechanisms*), I will give some historical examples. First, there are two quite simple cases. A plausible mechanism linking a clinical syndrome with a common cause is formulated. The features of this plausible mechanism are then specifically investigated. In this sort of (chronological replacement) case, the following might happen. Either an initial plausible mechanism is formulated that later turns out to be, roughly, correct, or that initial plausible mechanism turns out to be (mainly) mistaken. For instance, HSV was an highly plausible cause of cervical cancer because of the strong analogy between

[21] There are a few accounts of the role of analogy in theory construction. One notable example is Darden (1982), which is based on Hesse, 1963 and Oppenheimer, 1956. In short, Darden suggests four roles for analogy in theory construction: the exploitation of neutral analogy; deconstruction of causal links in new theory in terms of connections in similar theories; the construction of new theory from relevant pieces of old, unconnected theories and the recognition of shared abstraction between theories. This is of interest, yet these roles do not seem clearly circumscribed in my historical case. There is some other relevant work in the mechanisms literature too; (Darden, 2006; Craver, 2007; Bechtel, 2008). In any case, here I am trying to show that mechanisms need plausibility and analogy, and they are used in a fairly similar way to other, empirical, forms of evidence during the stages of mechanism construction.

[22] See also my short review of this area in section 1.2

cervical cancer and other tumours caused by herpesviruses.

These two patterns are fairly straightforward. But treating plausible mechanisms as simply a stage on the path to constructing a correct, empirically determined mechanism is to mis-read the way that things actually happen. It will fail to capture the many ways in which plausibility is important in causal inference and evaluation. Simply treating a plausible mechanism as an early stage of an empirical mechanism is to fail to notice that considerations of plausibility play a vital evidential role in making mechanisms. That it, a plausible mechanism is not simply a chronotropic stepping-stone on the path to a 'real' mechanism.

Instead, even mature mechanistic models with plenty of empirical support will tend to have had plausible statements used as evidence in their construction, may continue to contain sections that are primarily plausible (rather than empirical), and may be employed to do, say, interventions in a way that uses reasoning that is largely grounded in plausibility. This richness of interactions between mechanisms and plausibility shows itself in five plausible types of interaction between the success (or otherwise) of a causal claim and the helpful (or impedimential) role of either plausible or implausible mechanisms:

1. The construction or acceptance of a (correct) empirical mechanism is helped by its plausibility. Hopefully, this is not a controversial claim.

2. A (correct) empirical mechanism is helped by notions of its implausibility. This seems highly counter-intuitive, but it is a fairly common occurrence. For example, Burkitt's conviction that the BL tumour syndrome was in need of a novel causal explanation resulted from the implausibility of possible, alternative causal mechanisms that could be suggested. Similarly, once a mechanism was suggested that was capable of explaining the weird features of the tumour syndrome, it gained acceptance.

3. The acceptance of a correct empirical mechanism is impeded by its implausibility. This sort of case appears much more problematic, and sadly much more historically prominent, than 5. For instance, the notion that infection with *Helicobacter pylori* caused peptic ulcers was seen as highly implausible, yet one that was ultimately shown to be correct.

4. The acceptance of an incorrect empirical mechanism is helped by its plausibility. For example, HSV as a cause of cervical cancer.

5. The acceptance of an incorrect empirical mechanism is impeded by its implausibility. For example, McArdle's suggestion that the mechanism accounting for the clinical features of McArdle disease arose from the glucolytic pathway.

So the historical cases present a difficulty for the claim that mechanisms and plausible mechanisms are really the same thing. Plausible mechanisms, in place of empirical ones, look likely to make for bad causal claims. But yet plausibility still seems to play an important role in making and evaluating good causal claims.[23] In this section, I will attempt to address this tension. First, I will attempt to

[23] Dis-analogies and implausible phenomena are also evidentially also very useful, as I go on to discuss

differentiate and disambiguate plausibility and analogy. I will go on to support (from a metascientific stance) the notion that plausibility and analogy are capable of acting as extremely useful evidence for causal mechanisms, but are not an adequate replacement for a mechanism in a (RWT compatible) causal claim. I will do this by referring back to the earlier work on gappiness in mechanisms (in section 5.4.5), and by trying to suggest under which circumstances plausibility or analogy tend to be 'safe' ways of bridging gaps, and when they will tend to be more risky.

Anyway, the first step is to differentiate plausibility and analogy. Plausibility is general, and is relative to background knowledge. Thus claims about the behaviour of entities is said to be plausible if (unspecified other) entities tend to behave in similar ways. So, for example, if we are trying to build a plausible mechanism for the pattern of metastasis of a newly diagnosed tumour, we might suggest that it is likely that the smaller tumours found are likely to be metastases, while the largest tumour is likely to be the primary. This plausible claim would rely on (generalised) expressions of background knowledge about the behaviour of malignant cells. So theoretical claims about tumours in general, such as "*cancer cells spread from the primary tumour*" and "*the older the tumour, the larger the tumour*" and "*cancer cells tend to spread either by the blood, by the lymphatic system or by direct invasion*" would be (tacitly) employed in constructing and appraising such plausible claims. You do not give a description, in the main, of what specific bits of knowledge make a plausible mechanism really plausible.

Analogies, on the other hand, are specific. To make an analogy is, explicitly, to compare aspects of the causal system under examination with aspects of a second one, which will generally be rather better understood. Thus when we have a novel syndrome that resembles a second, known syndrome in its clinical features, we might (by analogy) suggest the operation of a similar causal mechanism. Alternatively, if we discover a new pathogen that shares many features with an existing agent with well-understood features, we might make an analogically supported claim about the likely effects of this, new, agent.[24] Again, like plausibility, this requires implicit consideration of background knowledge (e.g. "*viruses that encode similar proteins are likely to infect similar cells*"). But it also features the explicit consideration of one or more comparison systems. In other words, one has to specify what it is that the new situation is analogous *to*.[25]

Judgements about whether something is plausible, or whether two things are analogous, are unstable. Plausibility depends upon background knowledge in general. Analogy likewise depends upon changing knowledge about specifics. As my distinction is clearly based on the historical cases, and these cases occur over a fairly short period of time (roughly 1950-90), it seems possible that the senses of, and hence distinction between, the terms could also change. Thus I do not offer these distinctions as universal and unqualified definitions. They are, instead, relative to the historical cases in a fairly broad sense. With that in mind, how do plausibility and analogy enter mechanism construction? What roles do they play? First, evidence of both types have important heuristic roles early in causal inference. It seemed plausible that a phenomenologically consistent tumour syndrome had a (common) cause of some kind.

[24]This happens now with HPV. When a new type of the virus is discovered, sequence homology is used to determine which super-family it belongs to. This then guides the manner in which research is conducted on the new organism.

[25]This semantic difference, with the differential use of comparison, is the metascientific grounds for this distinction

Likewise, sexually transmitted diseases were noted to have analogous epidemiological features to cervical cancer, so perhaps they shared causal similarities too. In a related way, both plausibility and analogy played a methodological role at this early stage, in suggesting the experimental approaches suitable for understanding novel causal situations. I think these suggestions are fairly unproblematic, so I will not labour the point. But note, again, the different employment of the two terms in these sketched examples.

A more controversial claim for the role of plausibility and analogy is to negatively constrain the process of mechanism construction. For example, the apparently simultaneous, multi-focal nature of BL tumours appeared to be dis-analogous to the other known causes of jaw tumours (see footnote 15, page 159). This negative constraint excluded, or at least made less inviting, a range of possible causal hypotheses that could account for the BL syndrome. Similarly, the epidemiological analogies between venereal disease and cervical cancer ruled out a raft of ideas about chronic-degenerative cancer causation in this case. This process of ruling out competing causal hypotheses seems a vital part of causal domain-finding.

A third evidential role for plausibility and analogy follows from my suggestion, in section 5.4, that models and representations of mechanisms tend to be gappy. To reprise briefly: sometimes gaps are elective. So we might omit parts of a mechanistic model that are not relevant to our particular use of it. For example, if we are asked to make a prognostic judgement about a particular individual with cervical cancer, we will generally not use anything like a 'full' model or representation to do so. Instead, we will use an incomplete mechanistic model in which gaps will be bridged with plausible approximations of the behaviour of entities and activities. Of course, the degree of this ellipticism will depend on the intended purpose of the causal mechanism. So that is one sort of gappiness where plausibility is used, fairly safely, to make mechanisms operational.

But some sorts of gappiness are more troubling. Very often, sections of mechanisms are missing from every - rather than from a single - model. Perhaps this is because the relevant research has not yet been done for one reason or another. Or these sections of the mechanistic model might fall outside the specialist domains of biomedicine. Or perhaps, more worryingly, the relevant research proves impossible to complete successfully.[26] Obligate gaps like these are also bridged using plausible or analogical evidence. However, given that these these gaps sometimes arise because of real causal difficulties, these obligate gaps are likely to be more tricky to fill successfully than elective gaps. Just such a instance occurred with HSV and cervical cancer. There were significant, obligate, gaps in the causal models that related HSV infection with the development of cervical cancer. These were centred around the specific manner in which HSV caused (at a cellular-molecular level) cervical cancer. Yet these gaps were (easily) filled with plausible and analogical evidence. In general, herpesviruses were seen as plausible causes of cancer. A range of specific analogies (EBV and BL; *Herpes samirii*, *Herpes atales* and so on) were also used to fill in these details. This easy filling of a difficult gap masked the experimental difficulties encountered while attempting to unpick this causal mechanism. This means the experimental failures

[26] Again, this might be a feature of local incommensurability. Two different mechanistic models could be (and I think often are) constructed focusing on adjacent areas of the mechanism that cannot be happily linked together. This may well indicate a serious problem with the causal model.

failed to seriously challenge the overarching HSV-based causal model in the way that we might have wished them to.

So this is a good example of a dangerous bit of gap-filling. What makes a safe, versus what makes a dangerous, guess about the likely shape of a mechanism? Is there any clear way of differentiating the two? As a starting point, can we point to why the plausibility and positive analogies of HSV were so problematic in this case? I have one suggestion. In this instance, the manner of viral classification appears to be partly responsible. Analogies were made with other herpesviruses. However, these analogies appertained to the oncogenic activities of herpes viruses, both generally (herpes are a plausible cause of cervical cancer) and specifically (HSV is analogous to this specific oncogenic herpesvirus, or to that other one). Yet the classification used for herpesviruses, which is based on the location at which the various viruses spend their dormant phases and on genomic considerations (Cleator and Klapper, 2004b, 24) does not clearly relate to the properties of the virus that does the oncogenic causing. Now there are examples where classification is performed on features that are responsible for interesting bits of disease causation. For instance, enteric strains of *Escherichia coli* are partially classified based on their mechanisms of virulence (Picard et al., 1999). Analogical reasoning, based on these classifications, might well be 'safe'. But the HSV case was not so. Of course, in this specific instance we can be wise after the fact. But the general principle stands: we should think about how analogies are supposed to be similar.

So it might be as well to think carefully about what mechanisms make entities and activities similar if we wish to rely on the plausible. This approach might suggest a means of predicting which plausible or analogical evidence is likely to make for safe gap-filling, and which will make for risky gap-filling. In the case where both the causal antecedents, and descendants are known, analogies will tend to make safer guesses than plausibility. This requires that (unlike the HSV case) the behaviours or features that make the two analogous structures similar are essentially the same as the behaviours or features that are supposed to account for the mechanistic role of the (analogically supported) gap-filler.

In the case where the either antecedents or descendants are not well understood, though, plausible reasoning will tend to make for safer gap-filling than analogies. This is because, if the causal model that the analogical gap-filler fits into is not well-described, use of an inappropriate analogy may licence an inappropriately strong inference about the soundness of the general outlines of the whole causal model. As happened in the HSV case, this may lead to the neglect of better alternatives.

6.5 Evidence of difference-making

How does EDM participate in forming causal models? That is the simple version of the question I would like to address in this section. Here is the slogan version of the answer: EDM, relative to the functional specification of a mechanistically understood causal model, demonstrates that a particular aspect of a cause may bring certain effect(s) about in a specified population. This is a rather complicated slogan, admittedly. But actually the role of EDM is very easy to see in the historical cases. I will therefore begin with the examples, and I will draw out the details of my position on EDM using them.

Let us consider two typical, but high-quality, pieces of EDM. One is the large, prospective, case-control study that sought a link between EBV and BL (de-Thé et al., 1978). This work successfully

demonstrated the causal relationship between EBV and BL, and (as I argued earlier) was a vital part of this causal model becoming generally accepted by the medical community. The second example was also a large, prospective, case-control study. This is the Prague cervical cancer study (Vonka et al., 1984a,b; Krcmár et al., 1986) that attempted to find the link between HSV and cervical cancer. It, of course, failed to find such a correlation, and these negative results played a central role in bringing about the rejection of the HSV-cervical cancer theory by the research community. This pair of experiments are therefore very useful for exploring the nature of EDM, as they are capable of suggesting the conditions under which such (methodologically very similar) trials can act as either strong evidence for or against a particular causal model *tout court*.

We need EDM to tell us *that* mechanisms happen in a particular situation. We may, for example, formulate a very well-supported causal mechanism linking radon gas exposure to lung cancer using *in vitro* models. But it will be suitable evidence of difference-making that tells us whether this mechanism is really an important cause of lung cancer in a given population - by, say, showing that individuals in high radon areas really do suffer excessively from lung cancer that cannot be explained by the operation of other causal factors. In other words, this sort of evidence shows the net effect of a cause by giving the magnitude (in a non-critical sense) of the connection between cause and effect in a specified population. That anyway is how things should work.[27]

So, that is the manner in which difference-making participates in a causal model. But what characterises evidence of difference-making, in itself? That is, how do we gain knowledge about difference-making from data? What sorts of strategies tell us about difference-making in a given context? What makes it evidence of difference-making, rather than evidence of mechanism? First, being prospective (rather than retrospective) does not. Instead, it just makes it good quality evidence of difference-making. However, as an historical point, many useful bits of such evidence are methodologically similar. Prospective studies, after all, are very useful for minimising the likely effects of various systematic errors. This is particularly important for our definition of EDM, because it tends to be rather complex. In the historical cases, the mollification of various sources of bias were vital, because the (expected) effect sizes of the causal models were expected to be rather small. In both instances, something like a few tens of cases of each cancer were expected, from an enormous pool (tens of thousands) of starting subjects. Moreover, a large number of possible factors were recorded as being either factors to be controlled or investigated as possible causal factors. Many of these were either in themselves complicated (sexual behaviours) or difficult to clearly recall. Finally, the delay between the start of the study and the development of clinical disease was generally years, meaning that systematic errors like recall bias were likely to present a real problem. If the role of EDM is to figure out the overall effect of a cause, then in such causally complicated scenarios it has to arise from complex and subtle methodologies if it is to be useful and valid.

Similarly, the large number of trial participants, alone, does not make this EDM; it is simply the means by which an accurate estimation of the effect size could be conducted.[28] The McArdle case

[27] This is closely related to the distinction drawn by Hitchcock between net and component effects (Hitchcock, 2001).

[28] I might make a note here about statistical power. Interestingly, tumour viruses are very weak overall causes. For example,

is a clear counterexample to a necessary characterisation of EDM in terms of probabilities. However, because of the trend toward complexity in EDM, most useful EDM that is encountered in causal scenarios implicating a pathogen in the development of a clinical disease will tend to be probabilistic. Thus, as with ME and EM, we cannot give a simple methodological definition of EDM.

The types of entities under test are significant, though. In both cases, the trials examined the development of a clinically apparent disease syndrome (defined in terms of a mixture of symptoms, signs and investigation results) to serological evidence of viral infection.[29] Thus contrasts were sought between cases and controls over a range of features of the different causal mechanisms for development, or non-development of the disease. This does seem significant. In fact, I would like to suggest that a characteristic of EDM is that it looks for the relationship between entities that are of disparate types. This can take two forms. First, EDM can look for a correlation between entities that exist at widely differing levels of organisation. For example, it might be a correlation existing between a symptom (whole organism or organ-level) and a pathogen (cellular/macromolecular). Taking the (approximate) convention that causal mechanisms are represented with levels of organisation on the Y-axis, and the direction of mechanistic flow on the X, we might think of this as *vertical* separation. Alternatively, this might be *horizontal* separation, where the two phenomena under examination are separated by a large amount of intervening causal mechanism. This is a clear contrast with mechanistic evidence, which tends to concentrate on the activities of closely (time, space, type) related entities.

A second contrast can be drawn out from the relata of the evidence. Remembering that mechanistic evidence looks to demonstrate the manner in which entities interact via their activities, it seems possible to distinguish EDM by noting that it is concerned solely with the *existence* of entities. That is, EDM makes claims about entities only, and does not care (in detail) how they came about. It instead takes the form of existential claims about the things of the mechanism. That is not to say that EDM is static; quite the opposite. Instead, it tends to look for patterns of change in these entities, for contrasts between entities in different individuals, or changes that occur under different sets of background conditions. Thus EDM cannot arise from a simple reductionistic understanding of the types of entities causally responsible for a particular phenomena. Nor can it, alone, show how the cause operates. What it can do, when originated from, and granted context by, a particular mechanistic model, is to show how the entities that this mechanistic model describes, change overall. After all, if the BL prospective study had looked for evidence of infection with EBV *simpliciter* it would not have detected any causal link. Useful difference-making evidence has to demonstrate specific, rather than general, features of a mechanistic model.

There are things that these two studies do, evidentially, other than demonstrate the magnitude of the effect of the mechanism via a particular route. In the EBV case, they demonstrated the chronological

---

a maximum of 0.1% of those of the susceptible age, living in the lymphoma belt, with EBV infection, end up developing BL. Compare this to, say, the odds of developing clinically apparent measles given *de novo* infection with measles virus, which approach unity.

[29] The EBV study looked at a range of EBV antigens and antibodies, as well as a range of non-EBV serological indicators, while the HSV study looked at specific serological evidence of HSV2 infection

development of the condition, and the other differences in EBV seroconversion that obtained in patients versus control subjects. In the HSV case, the very careful controlling for age, age at first sexual intercourse, number of sexual partners, smoking and DKG[30] revealed no difference in HSV status between cervical cancer patients and controls matched for these factors. Not only did this suggest that the (plausible) HSV mechanism was faulty; it also suggested that, because these factors did make a profound overall difference to the rates of cervical cancer seen, that there was another pathogen at work. They also served to indicate the likely means of transmission (sexual) by which it would spread.

Both these more subtle outcomes and the main probability altering outcome are only intelligible as evidence for the existence and nature of a particular causal model when understood in concert with mechanism. They would be quite unintelligible without it. Furthermore, these (expensive and complicated) experiments simply would not have been conducted as they were in the absence of a postulated mechanism. Without such a mechanistic understanding of the disease, researchers simply would not have known which entities to look for.

In summary, it is the descriptive complexity of EDM that makes it non-mechanistic. EDM typically describes the relationship between participant entities which are widely separated in a given mechanism. We therefore typically expect EDM for medical mechanisms implicating a pathogen in the development of a disease to arise methodologically from population-level observational studies[31] But remembering that I claim that the business of causal inference is recursive, the mechanism for this causal situation will rely on other types of EDM, derived from research on the constituent behaviours of the mechanism. These (smaller-range and micro-level) EDMs are synthesised into the causal mechanism that describes the overarching mechanism implicating the pathogen in disease development. This is why I suggest that DME used in the construction of one mechanistic model can act as ME for the larger model. The judgement of evidence of difference-making versus of mechanism therefore depends upon the precise functional specification of the causal model in question. EDM for a causal model dealing with a small - entities, scope, narrower domain - causal model may well become EM in a larger model. Evidence that a particular causal route happens (smaller) becomes evidence for that causal route being capable of happening (larger).

I suggested that most EDM is probabilistic. It seems worthwhile exploring an instance when it was not. This comes from the McArdle case. Consider McArdle's investigation of George W. I suggest that the evoking of a range of disparate clinical phenomena (electrically silent cramps, lactate abnormalities, fatigue, muscle swelling) by exercise are the important EDM in this case. After all, this evidence consists of the description of a complex (chronological and spatial) relationship between fairly atomic entities. This then seems to be a species of DME. Why then was this high quality, useful, EDM? Why was there no need to perform a high-powered case-control study, say, for this to count as good evidence for causation? Why was it that limited comparisons between the metabolic functions of George W. and a small number of volunteer subjects sufficed? My suggestion is this. First, the appearance of the clinical features were

[30] Electrocautery of the transformational zone of the cervix. This was widely employed in the 1950s and 1960s as a means of managing precancerous conditions of the cervix (Vonka, personal communication).

[31] There are a number of legal and ethical difficulties with performing interventional studies in the field of tumour virology.

chronologically fairly immediate. This is most unlike the viral oncogenesis cases, in which a period of many years generally elapsed between infection and tumour development. Similarly, the effects of McArdle disease are rather simple and striking. This immediacy and easy observability mitigated the likely effects of the many types of systematic error to which observational studies are prone. Second, the cause of the clinical features (exercise) is fairly internally simple[32] and easy to describe. Thus it was not subject to the same vagaries of needing to give a clear, internal, mechanistic description as were the viruses. Finally, the clinical features described were similarly clear and unusual. While they were still in need of causal attribution, their phenomenological features were not.

6.6 Difference-making and the Reference Class Problem

As most EDM is probabilistic,[33] there is an important issue that must be addressed at this point. Say we wish to construct an RWT-type causal claim that deals with the manner in which infection with a particular virus leads to a particular tumour. The construction and evaluation of this causal claim will require that we generate or interpret appropriate EDM, such that we can accurately determine the probabilities associated with the instantiation of a particular mechanism in a population. How can we, from our probabilistic EDM determine the probability of the instantiation of this single causal-mechanistic route?

We might suggest that the probability of this particular event occurring should be the same as the general probability of events of the same type occurring. However, we can describe any particular event (or thing) in as many ways as we like, and we therefore have an infinite choice of classes of similar events to which we might assign the particular event. For example, we might classify a case of BL as a member of the malignant diseases, or a member of the diseases caused by EBV, or as a tropical disease, or as a common disease, or as a childhood disease and so on. But the probability of occurrence of each of these classes of disease will be widely differing. Both in practice and in theory, classification of an event to a single, objectively most appropriate reference class is not possible, and we can only give probabilities for single events that relate to their desciptions, rather than on the events themselves (Gillies, 2000, 119). Therefore, it may be impossible to give an objective probability for a single event. Reichenbach's formulation of this RCP is especially clear:

> If we are asked to find the probability holding for an individual future event, we must first incorporate the case in a suitable reference class. An individual thing or event may be incorporated in many reference classes, from which different probabilities will result. This ambiguity has been called the *problem of the reference class*. Assume that a case of illness can be characterized by its inclusion in the class of tuberculosis. If additional information is obtained from an X-ray, the same case may be incorporated into the class of serious cases of tuberculosis. Depending on the classification, different probabilities will result for the

[32] Although, the later discovery of the second wind phenomena revealed that it was perhaps not quite as simple as initially suspected

[33] Actually, I think a similar difficulty will also apply to non-probabilistic EDM too. For example, the EDM produced by use of case-control studies in the McArdle case will encounter similar descriptive difficulties.

prospective issue of the illness.

(Reichenbach, 1949, 374)

While this was originally described as a problem for frequentist theories of probability, Hájek (2007) has convincingly argued that slightly different versions of it will affect all informative theories of probability.[34] As the RWT certainly is intended to be an informative theory, albeit one that does not deal with probabilities exclusively, it too will have to provide some kind of solution to the RCP.

6.6.1 Salmon's solution(s) to the Reference Class Problem

While there may be no good theoretical solution to the RCP, by say modifying our theories of probability, there may be some tactics that can reduce the practical problems associated with it. In fact, Salmon suggested a number of such solutions. These are based around a simple rule for deciding which reference class we should assign our single event to. This is the *reference class rule*, and is as follows:

> ... choose the broadest homogeneous reference class to which the single event belongs.

(Salmon et al., 1971, 43)

Homogeneity requires some development. Salmon suggested two possible ways of formulating this term. His first suggestion was that homogeneity should be interpreted objectively:[35]

> To say that a reference class is homogeneous - objectively homogeneous for emphasis - means that there is no way, even in principle, to effect the relevant partition.

(Salmon, 1977, 399).

This still requires elaboration. Let us consider the situation where we are attempting to formulate a prognostic judgement about the probability that an individual with BL will survive for a particular period of time. We might begin by including BL in the class of all malignant disease. However, this reference class will clearly be non-homogeneous with respect to this probabilistic measure of survival. Some types of cancers kill almost all those affected very quickly, while other malignant diseases have very little effect on survival. Thus these more aggressive tumours (within this reference class) should have a very low probability of survival attached, while the rather slower tumours should have a higher probability. If we wish to make a accurate single case probabilistic prediction of survival for our patient with BL, we will

[34] *Informative* here is intended in the sense of those theories that can guide actions or inferences: 'The reference class problem besets those theories that are genuinely informative and that plausibly constrain our inductive reasonings and decisions' (Hájek, 2007, 564).

[35] In fact, Salmon persisted with objectively homogeneous reference classes, despite a number of difficulties. While I will go on to discuss Salmon's epistemic version of homogeneity later, it is important to note that Salmon did not regard this epistemic formulation as a suitable replacement for objective homogeneity: 'My first attempt at a detailed explication of objective homogeneity was badly flawed (1977b); an improved treatment, which I hope is more successful, was not published until 1984 (chap. 3). However that may be, I am still convinced that the concept of an *objectively homogeneous reference class* is legitimate and important. I maintain, for example, that the class of carbon-14 atoms is objectively homogeneous with respect to the attribute of spontaneous radioactive decay within 5730 years. If our concept does not fit cases of that sort, our explication must be at fault.' (Salmon, 1989, 171).

need to effect a partition in this reference-class such that the associated generic probability of survival is the same for all members of the class. We might try forming a partition of the class of malignant disease that just contains malignant disease of lymphoid cells. But, again, this reference class is distinctly non-homogeneous. A B-cell lymphoma is likely to take much longer to kill an affected individual than acute lymphoblastic leukaemia. So, again, we would have to partition this reference-class. When this process is repeated such that it comes to the point where there is no way, even in principle, to divide the reference class that would result in two meaningful partitions that differ in their associated probability measures, then we would have an homogeneous reference class.

While this sort of objective formulation of homogeneity may be useful in some contexts, it does not seem to be practically workable in the types of situations discussed in the historical cases. In fact, methodologies that work in this way are not a feature of these historical cases at all. This is because of a fairly simple difficulty that relates to the great number of possible causal factors which could, potentially, make a difference.

Consider a case in which a number, n, of binary variables are suspected to be possibly and plausibly causally relevant to the development of a disease. If we wish to do any single-case work with this situation, and if we also wish to follow an objective version of the reference class rule, then our difficulty will be as follows: for each n variables, we must form 2^n possible reference classes. Given that there are so many possible, plausible causal factors at play in any disease, this will almost certainly make our reference classes impossibly narrow. So following an objective formulation of the reference class rule seems an unworkable strategy here. However, Salmon also offers an alternative, epistemic formulation of the reference class rule, which might be more suitable in this case:

> When we know or suspect that a reference class is not homogeneous, but we do not know how to make any statistically relevant partition, we may say that the reference class is epistemically homogeneous.
>
> (Salmon et al., 1971, 44)

Now this, at the price of sacrificing the objective type of description of causality that interested Salmon, seems more promising. However, an exhaustive approach to partitioning here is, as with the objective formulation, not really workable either. Suppose we have about a dozen (dichotomous, binary) likely factors that might be causally significant. We might, for instance, take the range of risk factors that were identified during the early phases of research on cervical cancer.[36]

We will, in order to make sure that any reference class we assign a probability to is epistemically homogeneous, here have still to formulate more than 8000 partitions. Given the weak effects of most tumour viruses, this would require the creation of an enormous trial. In this example, the problem would be further exacerbated by the fact that most of these variables have a wide range of values that might, plausibly, lead to different effects. Further, some of these variables are even more complicated. For instance, both youth and age are, broadly, risk factors for this disease. So actually creating an

[36] To recap from section 3.2, these were sex, marriage, parity, faith, youth, syphilis, social class, race, age, male partner circumcision, smoking and use of oral contraceptives

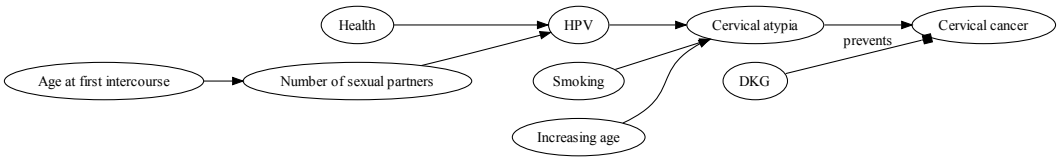

Figure 6.1: Cervical cancer example. Here, for the sake of clarity, a number of features have been omitted. Most importantly, at each stage along the health–HPV–cervical atypia–cervical cancer track, side-branches that express either i: the non-progression of the relevant state or ii: a regression to a prior state have been left out. Thus only the important, pathogenic tracks of this branched mechanism are shown.

exhaustive set of data partitions does not seem like an operationally effective strategy for understanding the causation of cervical cancer. Is there a more effective methodology that we might use to fulfill the epistemic version?

6.6.2 Mechanistic models and the reference-class problem

Mechanistic models might provide a principled means of fulfilling the epistemic formulation of the reference class rule. There also seems some historical support for this claim. For instance, the Prague cervical cancer study (Vonka et al., 1984a; Vonka et al., 1984b; Krcmár et al., 1986) neither attempted to match controls over a totally exhaustive set of causal factors (as we might if we were following the objective formulation of the reference class rule), or even an exhaustive set of all identified possible causal factors (c.f. an exhaustive epistemic interpretation of it). Instead, researchers performed case-matching on the basis of just five factors. These five factors - age, age at first sexual intercourse, number of sexual partners, smoking and history of DKG - were seen, at the time, as being useful indicators of all those possible factors likely to make some sort of causal difference (Vonka, personal communication). Now thinking about just five factors has major advantages. Not least, if we interpret them dichotamously, we have only 32 different partitions to make, rather than many thousands. And even with this highly restricted class of variables used to define reference classes, the study that used these factors provided an highly effective demonstration of the falsity of the HSV theory.

I think this sort of example provides us with a principled means of selecting those features which we might employ to satisfy the epistemic formulation of the reference class rule. Models of mechanisms provide the means by which we can do this. In more detail, it is the point at which the pathological routes of our mechanistic model branching off from physiology, or their direct descendants, that are ideal targets for epistemically dealing with the RCP. These branching points will, as earlier suggested, also generally arise from setting effects. Depending upon the ease with which we can collect data about them, any one of a range of possible features around this type of branching point will be a good classificatory feature.

As an example, the features identified in the Prague study might be used to construct a schematic mechanistic model of the relevant aetiological features (figure 6.1).

Here, only those activities and entities that constitute cervical cancer pathogenesis are shown. This pathogenic mechanisms is influenced by a number of likely causal factors. The five factors (age at first intercourse, number of sexual partners, increasing age, smoking status and history of DKG) are related together in a plausible manner by a mechanistic model. If we are interested in, say, the probability of a particular individual developing cervical cancer we can use such a non-exhaustive approach to provide a principled means of resolving the reference-class problem in practice. We can, when attempting to partition our EDM, refer to this example mechanism as a way of deciding which partitions we need to form. In general, when we are trying to understand the occurrence of a particular effect, we will be interested the various parents of that effect, as well as any other effects these parents have, and the other parents these effects have.[37] In this toy example, we only have a single effect of interest. However, as much recent research on cervical cancer has involved the identification of disease markers, there might well develop a case such that a number of effects, and their causes, would participate in deciding just how to partition statistical data in a way that does not require exhaustive consideration of all possible factors.

In summary, as the RCP forces us to restrict the features that we examine when we collect EDM, and the rational solution to this restriction is to employ a mechanistic model to permit a principled selection of suitable features, then we can explain why the RWT is a good descriptive theory of causality in medicine. This suggestion also lends support to the normative interpretation of the RWT that I will explore in the next section.

[37] i.e. we will be interested in the Markov Blanket (Pearl, 1987, 249). However, I avoid this term because of the very significant connection that it has to Bayesian networks. While these tools would provide an extremely useful means of dealing with this type of problem specifically, I will not develop this account in such a way here.

Chapter 7

Conclusion: The Russo-Willamson Thesis Redux

I have argued that the RWT is a good descriptive theory. Chapters 2 and 3 made the historical case for it, while chapters 5 and 6 made the related methodological case. However, I think the RWT should also be interpreted normatively. This normative aspiration gains support from both aspects of the descriptive adequacy of the theory. But I think the RWT is capable of telling us what is, and what is not, an high-quality causal claim in medicine.

In what follows, I therefore give some suggestions about how this might be achieved. This begins, in section 7.1, with a clear statement of what RWT-type causal models are, why they work, and how they are related to the generation of appropriate evidence. This, then, explains how the descriptive methodological case supports a normative interpretation. Then, in section 7.2, I will give a clear restatement of the descriptive historical case, and how this also supports a normative interpretation. Finally, I'll conclude with some remarks on why we might want to employ a normative interpretation of the RWT.

7.1 The Russo-Williamson Thesis redux

RWT-type causal models are composed of integrated evidence of mechanism and difference making. Evidence of mechanism is integrated into a mechanistic model, while evidence of difference-making (generally as probabilities of the instantiation of various parts of this mechanistic model) is put together in a way that examines particular features of this mechanistic model. Thus, RWT-causal models are highly context-dependent, epistemic, devices for understanding causal relationships that depend upon two interdependent types of evidence.

Both types of constitutive evidence are necessary to support a high-quality causal claim. Mechanistic models will be defeated by unfaithful causal structures, and by causal irrelevance, while systems of EDM alone will be defeated by confounding variables, while the generation of useful EDM is often implausible in the absence of some kind of mechanistic model. Either type of evidence-synthesis alone is inadequate.

As an illustration, consider a purely hypothetical research programme looking to discover the relationship, if any, between smoking and atherosclerosis-induced heart disease in a population. A mecha-

nistic model may be developed that suggests that smoking directly causes atherosclerosis. However, this mechanistic model may not accurately reflect what actually occurs in our population of interest. This may be as a result of two different problems with solo mechanistic models. First, the structure of our (ontic) mechanism in the population may be unfaithful. For instance, smoking may prevent weight gain. As excessive weight is also cause of atherosclerosis, and therefore heart disease, the weight-reducing function of smoking could outweigh the direct atherogenic effects, leading to an overall decrease in heart disease in smokers. Second, our mechanistic model may be irrelevant to the actual (ontic) disease mechanisms that occur in our population. It may be the case that the evidence which we use to construct our mechanistic model poorly reflects the situation in the relevant population. For instance, smoking may be unknown in the actual population. For another, the mechanistic model concerning smoking might solely concern itself with very heavy smoking, whereas our population might contain exclusively light smokers.

To resolve these difficulties, we need evidence of what the net effect of smoking actually is on the population. This is evidence of difference-making. Alone, this too is inadequate. First, our EDM may be unreliable because of confounding variables. Smoking may correlate with some third variable that also causes heart disease, such as alcoholism. If we do not know that some smokers are also alcoholics, we may overestimate the deleterious effects of smoking. Knowledge of mechanistic models allows us to control for such correlates and thereby to demonstrate accurately the overall direct effects of smoking. Without a mechanistic model, the reference class problem makes it impossible to attempt to control for all possible plausible causal factors. Second, useful EDM may be methodologically secondary to a mechanistic model. It may well be the case that we cannot generate the kinds of EDM that could usefully participate in finding out about the operation of our cause without understanding the relevant mechanism of disease. In our hypothetical case, we may never think to study the effects of heavy versus light smoking on the development of heart disease. Similar, stronger, examples of this methodological sequence are to be found in the EBV case (c.f. de-Thé et al., 1978), and in the HCC case (c.f. vaccine production). In the EBV case, it seems profoundly implausible to claim that the decision to look for the very specific serological factors in the BL-EBV prospective study would have been taken without an understanding of how EBV interacts with the immune system. Similarly, in the HCC case, I do not think we can plausibly claim that the development of two different, highly effective, vaccines came about without understanding, in some detail, the nature of the relevant aetiological agent, the manner in which it infects cells, the normal way in which the immune system functions and can be modified, and so on. Without a detailed mechanistic understanding of these situations, the decision to perform such detailed observational or interventionist studies would be nothing short of miraculous.

Put together, these two types of evidence show us the overall effect of a given causal factor in a given population, while also detailing the contributions of each branch of the causal mechanism linking cause and effect. A RWT-causal model therefore presents us with a causal road-map for a particular factor in a population. The mechanistic aspect tells us which causal routes can happen, and describes their morphology, while the difference-making aspect tells us which causal routes are actually taken, and

under which circumstances.

Thinking of RWT-causal models as causal road-maps allow the understanding of a number of issues regarding causality that are generally a bit difficult to understand. First, causation by omission should be construed as being a change in the actual causal route taken through an RWT-type causal structure. It is not the case that a negative cause represents termination of a causal mechanism. Rather, it is analogous to the blocking of a street that is conventionally open to traffic. This leads to a rerouting of causal instances around this blockage via alternative routes. A description of these routes, in terms of an altered RWT-type causal model, is how causation by omission should be understood.

Second, these RWT-type causal structures deal with generic causality. However, because of their treatment of the reference class problem, understanding and predicting the operation of a single-case cause simply requires the description of the possible routes that could be taken in the particular instance.[1]

Third, the two sorts of effects outlined in section 5.7 are similarly easy to understand. Sequence effects can be thought of as causal routes that move through the causal landscape without changing its arrangement. Setting effects, on the other hand, are causal routes that change the shape of the landscape, or the probability of subsequent instances taking particular causal routes.

Finally, the very important issue of finding effective strategies[2] is greatly simplified by using road-maps. Effective points for intervention will generally be those that alter branching points that arise by setting effects. We can either do this by altering the setting effect itself, instantiating an alternative setting effect, or by changing other factors such that altered sequence effects bypass the deleterious effects of the pathological setting effect.

7.2 Integration of evidence in practice and the methodologies of bioscientific research programmes

RWT-type causal models are highly pluralistic with respect to evidence. No particular experimental methodology generates evidence that is either necessary or sufficient to unequivocally support a causal claim. Instead, a wide range of different types of evidence participates in sustaining and developing a causal model.

The change in the morphology of these causal models changes as their associated research programme changes. This close relationship between a developing causal claim, and the organisation of research, is clearly visible in the historical cases. For instance, the initial stage of the BL-EBV research programme consisted primarily of formulating a basic clinical and epidemiological description that appeared to arise from the operation of some kind of anomalous, novel causal process. However, once this disease was demonstrated to be a malignant one with a characteristic geographical distribution, a change was effected in the kinds of evidence that were considered of importance in determining the nature of causality in this case. This too was reflected by a change in the patterns of personal and institutional involvement in the case, with the recruitment of various experts in both malignant and infectious

[1] c.f. also Russo and Williamson's treatment of single-case and generic causation (Russo and Williamson, 2010).

[2] c.f. **RW2**

diseases.

This initial investigation of BL as an infectious tumour underwent a further organisational change when the theory that it might be caused by a virus became plausible. This phase, when EBV was discovered in the tumour cells, and investigated as the causal agent for BL was again characterised by a change in the types of individuals and institutions involved. It was also characterised by the adoption, and in some cases development, of a novel range of experimental methodologies. Finally, research looked to demonstrate the applicability of the EBV mechanism to disease populations. This involved constituting seroepidemiological research in the field that relied on this newly demonstrated mechanism. This phase also involved research into a number of (clinically and aetiologically) related disease states.

This change in the organisation of research around a developing causal model is characteristic of successful research programmes. More specifically, research programmes that are likely to produce useful causal models are those in which various research groups and experimental methodologies become conjoined around (local) causal models. This is initially triggered by the search for a particular causal fact (aetiology, therapy) and is shaped by the interpretive framework adopted (cf. the distinctly oncovirological research programme that grew up around EBV). Research programmes become conjoined by the integration of laboratory and epidemiological investigation. This integration is methodologically recursive. At some points, questions raised by laboratory work guides epidemiology, while at other points it will be the case that epidemiological research guides laboratory workers. Research in different domains become conjoined in causally successful research programmes (such as the BL-EBV, HPV-cervical cancer programmes), but such conjoining does not happen in unsuccessful ones (such as the HSV-cervical cancer programme).

Such development of a conjoint research programme leads to the production of an epistemic web of different forms of evidence, an interdependent arrangement of the diverse types of EM and EDM that can be employed in the construction of RWT-type causal models. This web of evidence arises from the bidirectional interplay between types of methodologies that can only come about when multidisciplinary research programmes are so organised. However, causally interpreting such diverse evidence requires an interlevel understanding of causality. As Machamer et al. (2000, 23) suggests, 'It is the integration of different levels into productive relations that renders the phenomenon intelligible...' While this suggestion relates to mechanisms as a means of understanding causality, I think that this is better construed as a claim about the interlevel and interdisciplinary advantages of RWT-type causal models. After all, those who would reason exclusively with mechanisms cannot account for the role of different types of evidence in supporting causal claims.

7.3 Final conclusions

Medicine, argues Russo and Williamson, seeks causes for cognitive and action-oriented reasons. A major advantage of the normative RWT is that it is capable of discriminating between strong and weak causal knowledge in both these contexts.

An RWT-type causal model is not an absolute prerequisite for all forms of causal knowledge. We may discover that, say, a particular chemical produces some therapeutic effect in the absence of a mech-

anistic model capable of explaining its effects. There are numerous examples of such discoveries: some of the therapeutic effects of aspirin, anaesthetics and so on were discovered without an understanding of their mechanisms of action. However, the causal knowledge that we had about these treatments was limited. While we knew that some of their effects occurred, we could not speak to why. Nor, in several cases, did we know of a whole range of undesirable effects that they also possessed.

For example, while the understanding that aspirin could be used as an analgesic agent dates from the 1890s. This discovery did not depend upon a mechanistic model, nor an RWT-type causal model, of its mechanism of action. And aspirin was used, very successfully, with this limited causal knowledge for many years. However, the discovery of its mechanism of action, and the later understanding of its wide range of adverse effects (discussed in section 5.3.3) did arise from such investigations. And, as a result, it is now the case that aspirin is no longer prescribed for analgesia, as this research has, by a combination of clinical and laboratory research, demonstrated that a wide range of alternative agents are safer and more effective. Furthermore, other apparently valuable features of this drug that were previously discovered are also being questioned as clinical research becomes more closely integrated with the mechanistic understanding of the drug. For example, it has recently been suggested that the use of aspirin as a primary preventive agent against vascular disease is much less effective than previously thought (Baigent et al., 2009). All this means that we should not necessarily discard all causal knowledge of this type. What it does mean is that it should be employed in a more circumspect fashion. The RWT is suited to exactly this purpose: causal claims that are supported by appropriate evidence of both mechanism and difference-making are very likely to be more reliable than those that are not. Use of this demanding causal methodology is therefore an effective means by which to critically examine causal claims in medicine, and to suggest in which circumstances a particular causal claim is likely to be an effective guide to practice, and when it is not. This is extremely valuable, as poorly understood causal claims may well survive, unchallenged, in medical practice and cause great harm.

Bibliography

Abboud, H.E., 1993. "Growth Factors in Glomerulonephritis." *Kidney International*, **43**(1): 252–67. 18

Adam, E., Kaufman, R.H., Adler-Storthz, K., Melnick, J.L., and Dreesman, G.R., 1985. "A Prospective Study of Association of Herpes Simplex Virus and Human Papillomavirus Infection with Cervical Neoplasia in Women Exposed to Diethylstilbestrol *in Utero*." *International Journal of Cancer*, **35**(1): 19–26. 65

Adam, E., Kaufman, R.H., Melnick, J.L., Levy, A.H., and Rawls, W.E., 1973. "Seroepidemiologic Studies of Herpesvirus Type 2 and Carcinoma of the Cervix. IV. Dysplasia and Carcinoma *in situ*." *American Journal of Epidemiology*, **98**(2): 77–87. 64

Aitken-Swan, J. and Baird, D., 1965. "Circumcision and Cancer of the Cervix." *British Journal of Cancer*, **19**(2): 217–27. 59

———, 1966a. "Cancer of the Uterine Cervix in Aberdeenshire. Aetiological Aspects." *British Journal of Cancer*, **20**(4): 642–59. 56

———, 1966b. "Cancer of the Uterine Cervix in Aberdeenshire. Epidemiological Aspects." *British Journal of Cancer*, **20**(4): 624–41. 56, 57, 58

Alexander, E., 1973. "Possible Etiologies of Cancer of the Cervix Other Than Herpesvirus." *Cancer Research*, **33**(6): 1485–90. 60, 61

Allday, M.J. and Crawford, D.H., 1988. "Role of Epithelium in EBV Persistence and Pathogenesis of B-Cell Tumours." *The Lancet*, **331**(8590): 855–7. 27

Almeida, J., Howatson, A., and Williams, M., 1962. "Electron Microscope Study of Human Warts; Sites of Virus Production and Nature of the Inclusion Bodies." *The Journal of Investigative Dermatology*, **38**(4): 337–45. 66

Almeida, J., Oriel, J., and Stannard, L., 1969. "Characterization of the Virus found in Human Genital Warts." *Microbios*, **3**: 225–232. 67, 69

American Psychiatric Association, 1994. *DSM-IV-TR: Diagnostic and Statistical Manual of Mental Disorders*. Washington DC: American Psychiatric Press Inc., 4th Revised edition. 156

Andervont, H.B. and Bryan, W.R., 1944. "Properties of the Mouse Mammary-tumor Agent." *Journal of the National Cancer Institute*, **5**: 143–9. 16

Andrae, B., Kemetli, L., Sparen, P., Silfverdal, L., Strander, B., Ryd, W., Dillner, J., and Tornberg, S., 2008. "Screening-Preventable Cervical Cancer Risks: Evidence From a Nationwide Audit in Sweden." *Journal of the National Cancer Institute*, **100**(9): 622–9. 54

Anonymous, 1919. "The Genesis of Cancer of the Uterus." *The Lancet*, **194**(5017): 742–3. 56

———, 1947. "Homologous Serum Hepatitis." *The Lancet*, **250**(6480): 691–2. 101

———, 1975. "The Nobel Prize in Physiology or Medicine 1975." URL `http://nobelprize.org/nobel/prizes/medicine/laureates/1975/`. 16

———, 1981. "Human Virus, Hepatic Cancer." *The Lancet*, **318**(8260/1): 1394–5. 105

———, 1989. "way, n.1." In *The Oxford English Dictionary. 2nd ed. OED Online.*, Oxford University Press. URL `http://dictionary.oed.com/cgi/entry/50281826`. 139

———, 2008. "Cheaper HPV Vaccines Needed." *The Lancet*, **371**(9625): 1638. 75

Applegarth, D.A., Toone, J.R., and Lowry, R.B., 2000. "Incidence of Inborn Errors of Metabolism in British Columbia, 1969–1996." *Pediatrics*, **105**: e10. 95

Aronowitz, R., 2004. "Review: The Rise of Causal Concepts of Disease: Case Histories. Codell Carter, K." *International Journal of Epidemiology*, **33**(4): 911–2. 18

Aronson, J.L., 1971. "On the Grammar of 'Cause'." *Synthese*, **22**(3/4): 414–30. 146

Aurelian, L., Strandberg, J.D., Melendez, L.V., and Johnson, L.A., 1971. "Herpesvirus Type 2 Isolated from Cervical Tumor Cells Grown in Tissue Culture." *Science*, **174**(4010): 704–7. 63

Baigent, C., Blackwell, L., Collins, R., Emberson, J., Godwin, J., Peto, R., Buring, J., Hennekens, C., Kearney, P., Meade, T., Patrono, C., Roncaglioni, M.C., and Zanchetti, A., 2009. "Aspirin in the Primary and Secondary Prevention of Vascular Disease: Collaborative Meta-Analysis of Individual Participant Data from Randomised Trials." *The Lancet*, **373**(9678): 1849–60. 183

Baltimore, D., 1970. "RNA Dependent DNA Polymerase in Virions of RNA Tumour Viruses." *Nature*, **226**(5252): 1209–11. 16

Band, V., Caprio, J.D., Delmolino, L., Kulesa, V., and Sager, R., 1991. "Loss of p53 Protein in Human Papillomavirus Type 16 E6-Immortalized Human Mammary Epithelial Cells." *Journal of Virology*, **65**(12): 6671–6. 73

Bang, F.B., 1952. "Cellular Pathology of Virus Infections as Seen with the Electron Microscope." *Annals of the New York Academy of Sciences*, **54**(Viruses as Causative Agents in Cancer): 892–901. 15

Bauer, H.M., Ting, Y., Greer, C.E., Chambers, J.C., Tashiro, C.J., Chimera, J., Reingold, A., and Manos, M.M., 1991. "Genital Human Papillomavirus Infection in Female University Students as Determined by a PCR-Based Method." *Journal of the American Medical Association*, **265**(4): 472–7. 74

Beasley, R.P., Lin, C., Hwang, L., and Chien, C., 1981. "Hepatocellular Carcinoma and Hepatitis B Virus: A Prospective Study of 22 707 Men in Taiwan." *The Lancet*, **318**(8256): 1129–33. 103

Bechtel, W., 1994. "Levels of Description and Explanation in Cognitive Science." *Minds and Machines*, **4**(1): 1–25. 124, 137

———, 2001. "The Compatibility of Complex Systems and Reduction: A Case Analysis of Memory Research." *Minds and Machines*, **11**(4): 483–502. 123

———, 2008. *Discovering Cell Mechanisms: The Creation of Modern Cell Biology*. Cambridge: Cambridge University Press. 123, 161, 166

Bechtel, W. and Abrahamsen, A., 2005. "Explanation: A Mechanist Alternative." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, **36**(2): 421–41. 24, 118, 122, 123

Bechtel, W. and Richardson, R.C., 1993. *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Princeton, NJ: Princeton University Press. 122

Beebee, H., 2004. "Causing and Nothingness." In J. Collins, N. Hall, and L.A. Paul, editors, *Causation and Counterfactuals*, Cambridge, MA: MIT Press. 291–308. 146, 147

Bell, T.M., Massie, A., Ross, M.G., Simpson, D.I., and Griffin, E., 1966. "Further Isolations of Reovirus Type 3 from Cases of Burkitt's Lymphoma." *British Medical Journal*, **1**(5502): 1514–7. 47

Bell, T.M., Massie, A., Ross, M.G., and Williams, M.C., 1964. "Isolation of a Reovirus from a Case of Burkitt's Lymphoma." *British Medical Journal*, **1**(5392): 1212–3. 47

Biberstein, H., 1944. "Immunization Therapy of Warts." *Archives of Dermatology and Syphilology*, **50**(1): 12–22. 66

Biggs, P.M., 2001. "The History and Biology of Marek's Disease." In K. Hirai, editor, *Marek's Disease*, Berlin: Springer. 1–24. 15

Bittner, J.J., 1936. "Some Possible Effects of Nursing on the Mammary Gland Tumor Incidence in Mice." *Science*, **84**(2172): 162. 16

Bivins, J.A., 1953. "The Growth in the Developing Chicken Embryo of a Filtrable Agent from Verruca Vulgaris." *The Journal of Investigative Dermatology*, **20**(6): 471–81. 67

Black, P., Hartley, J., Rowe, W., and Huebner, R., 1963. "Transformation of Bovine Tissue Culture Cells by Bovine Papilloma Virus." *Nature*, **199**(4897): 1016–8. 68

Bluhm, R., 2005. "From Hierarchy to Network: A Richer View of Evidence for Evidence-Based Medicine." *Perspectives in Biology and Medicine*, **48**(4): 535–47. 85

Blumberg, B.S., Alter, H.J., and Visnich, S., 1965. "A "New" Antigen in Leukemia Sera." *Journal of the American Medical Association*, **191**(7): 541–6. 101

Blumberg, B.S., Sutnick, A.I., and London, W.T., 1968. "Hepatitis and Leukemia: their Relation to Australia Antigen." *Bulletin of the New York Academy of Medicine*, **44**(12): 1566–86. 101, 102

Boiron, M., Leby, J., Thomas, M., Friedmann, J., and Bernard, J., 1964. "Some Properties of Bovine Papilloma Virus." *Nature*, **201**(4917): 423–4. 68

van den Bosch, C., 2004. "Is Endemic Burkitt's Lymphoma an Alliance Between Three Infections and a Tumour Promoter?" *The Lancet Oncology*, **5**(12): 738–46. 48

Bosch, F.X., Lorincz, A., Muñoz, N., Meijer, C.J.L.M., and Shah, K.V., 2002. "The Causal Relation Between Human Papillomavirus and Cervical Cancer." *Journal of Clinical Pathology*, **55**(4): 244–65. 54, 63

Bosch, F.X., Manos, M.M., Muñoz, N., Sherman, M., Jansen, A.M., Peto, J., Schiffman, M.H., Moreno, V., Kurman, R., and Shah, K.V., 1995. "Prevalence of Human Papillomavirus in Cervical Cancer: A Worldwide Perspective. International Biological Study on Cervical Cancer (IBSCC) Study Group." *Journal of the National Cancer Institute*, **87**(11): 796–802. 74

Bosch, F.X., Muñoz, N., de Sanjosé, S., Izarzugaza, I., Gili, M., Viladiu, P., Tormo, M.J., Moreo, P., Ascunce, N., and Gonzalez, L.C., 1992. "Risk Factors for Cervical Cancer in Colombia and Spain." *International Journal of Cancer*, **52**(5): 750–8. 74

Boshart, M., Gissmann, L., Ikenberg, H., Kleinheinz, A., Scheurlen, W., and zur Hausen, H., 1984. "A New Type of Papillomavirus DNA, Its Presence in Genital Cancer Biopsies and in Cell Lines Derived from Cervical Cancer." *The EMBO Journal*, **3**(5): 1151–7. 71, 74

Bouchier, I.A.D., Ellis, H., Fleming, P.R., and French, H., 1996. *French's Index of Differential Diagnosis*. Oxford: Butterworth Heinemann, 13th edition. 23, 159

Bouvier, G.L., Sussman, M., and Crawford, L., 1966. "Antigenic Diversity of Mammalian Papillomaviruses." *Journal of General Microbiology*, **45**(3): 497–501. 69

Bowesman, C., 1960. *Surgery and Clinical Pathology in the Tropics*. Edinburgh: E & S Livingstone. 35, 37

Boyd, A.L. and Orme, T.W., 1975. "Transformation of Mouse Cells After Infection with Ultraviolet Irradiation-Inactivated Herpes Simplex Virus Type 2." *International Journal of Cancer*, **16**(4): 526–38. 62

Boyd, J. and Doll, R., 1964. "A Study of the Aetiology of Carcinoma of the Cervix Uteri." *British Journal of Cancer*, **18**(3): 419–34. 56, 64

Breitburd, F., Salmon, J., and Orth, G., 1997. "The Rabbit Viral Skin Papillomas and Carcinomas: A Model for the Immunogenetics of HPV-Associated Carcinogenesis." *Clinics in Dermatology*, **15**(2): 237–47. 68

Brew, D. and Jackson, J.G., 1961. "Lymphosarcoma in the Ovary in Young African Girls in Nigeria." *British Journal of Cancer*, **14**(4): 621–6. 35

Broadbent, A., 2010. "Inferring Causation in Epidemiology: Mechanisms, Black Boxes and Contrasts." In P. McKay Illari, F. Russo, and J. Williamson, editors, *Causality in the Sciences*, Oxford: Oxford University Press. Forthcoming. 86, 87

Brown, J.B. and O'Keefe, C.O., 1928. "Sarcoma of the Ovary with Unusual Oral Metastases." *Annals of Surgery*, **87**(3): 467–71. 38

Brown, M.S. and Goldstein, J.L., 1992. "Koch's Postulates for Cholesterol." *Cell*, **71**(2): 187–8. 18

Bunting, H., 1953. "Close-Packed Array of Virus-Like Particles Within Cells of a Human Skin Papilloma." *Proceedings of the Society for Experimental Biology and Medicine*, **84**(2): 327–32. 66

Burford, W.H., Ackerman, L.V., and Robinson, H.B.G., 1944. "Symposium on Twenty Cases of Benign and Malignant Lesions of the Oral Cavity, from the Ellis Fischel State Cancer Hospital, Columbia, Missouri: Case 1: Lymphosarcoma in a Child." *American Journal of Orthodontics and Oral Surgery*, **30**(7): 353–5. 38

Burkitt, D.P., 1951. "Primary Hydrocele and Its Treatment. Review of Two Hundred Cases." *The Lancet*, **257**(6669): 1341–3. 38

———, 1958. "A Sarcoma Involving the Jaws in African Children." *The British Journal of Surgery*, **46**(197): 218–23. 28, 29

———, 1961. "Observations on the Geography of Malignant Lymphoma." *East African Medical Journal*, **38**(11): 511–4. 38, 39, 40

———, 1962a. "A Children's Cancer Dependent on Climatic Factors." *Nature*, **194**(4825): 232–4. 34, 37, 38

———, 1962b. "A Children's Cancer with Geographical Limitations." In R.W. Raven, editor, *Cancer Progress*, London: Butterworth. 102–3. 34

———, 1962c. "Determining the Climatic Limitations of a Children's Cancer Common in Africa." *British Medical Journal*, **2**(5311): 1019–23. 34

———, 1962d. "A Lymphoma Syndrome in African Children." *Annals of the Royal College of Surgeons of England*, **30**(4): 211–9. 34, 37

———, 1962e. "A 'Tumour Safari' in East and Central Africa." *British Journal of Cancer*, **16**(3): 379–86. 34, 40, 41

———, 1962f. "A Tumour Syndrome Affecting Children in Tropical Africa." *Postgraduate Medical Journal*, **38**(436): 71–9. 34, 37

———, 1983. "The Discovery of Burkitt's Lymphoma." *Cancer*, **51**(10): 1777–86. 30, 31, 32, 33, 155, 159

Burkitt, D.P. and Davies, J.N.P., 1961. "Lymphoma Syndrome in Uganda and Tropical Africa." *Medical Press*, **245**: 367–9. 35, 40

Burkitt, D.P., Nelson, C.L., and Williams, E.H., 1963. "Some Geographical Variations in Disease Pattern in East and Central Africa." *East African Medical Journal*, **40**(1): 1–6. 31

Burkitt, D.P. and O'Conor, G.T., 1961. "Malignant Lymphoma in African Children. I. A Clinical Syndrome." *Cancer*, **14**(2): 258–69. 33, 36

Burkitt, D.P. and Wright, D.H., 1963. "A Lymphoma Syndrome in Tropical Africa with a Note on Histology, Cytology, and Histochemistry." *International Review of Experimental Pathology*, **2**: 67–138. 37, 38, 39, 40

Butel, J.S., 1972. "Studies with Human Papilloma Virus Modeled After Known Papovavirus Systems." *Journal of the National Cancer Institute*, **48**(2): 285–99. 67

Buxton, P.A., 1955. *The Natural History of Tsetse Flies: An Account of the Biology of the Genus* Glossina *(Diptera)*. London: H.K. Lewis. 40

Buynak, E.B., Roehm, R.R., Tytell, A.A., Bertland, A.U., Lampson, G.P., and Hilleman, M.R., 1976. "Vaccine Against Human Hepatitis B." *Journal of the American Medical Association*, **235**(26): 2832–4. 104

Callahan, R. and Smith, G., 2008. "The Mouse as a Model for Mammary Tumorigenesis: History and Current Aspects." *Journal of Mammary Gland Biology and Neoplasia*, **13**(3): 269. 16

Capponi, M., 1953. "Note Sur Le Cancer Au Cameroun [Note on Cancer in the Cameroons]." *Bulletin de la Société de Pathologie Exotique*, **46**(4): 605–11. 34

Carter, K.C., 1987. *Essays of Robert Koch*. Westport, CT: Greenwood Press. 18

———, 2003. *The Rise of Causal Concepts of Disease: Case Histories*. Hants: Ashgate. 18, 93, 99

Cartwright, N., 1994. *Nature's Capacities and Their Measurement*. Oxford: Clarendon Press. 132

Cassai, E., Rotola, A., Meneguzzi, G., Milanesi, G., Garsia, S., Remotti, G., and Rizzi, G., 1981. "Herpes Simplex Virus and Human Cancer. I. Relationship Between Human Cervical Tumours and Herpes Simplex Type 2." *European Journal of Cancer*, **17**(6): 685–93. 65

Cassel, J., 1976. "The Contribution of the Social Environment to Host Resistance: The Fourth Wade Hampton Frost Lecture." *American Journal of Epidemiology*, **104**(2): 107–23. 19

Catalano, L.W. and Johnson, L.D., 1971. "Herpesvirus Antibody and Carcinoma *in situ* of the Cervix." *Journal of the American Medical Association*, **217**(4): 447–50. 64

Chakraborty, P.R., Ruiz-Opazo, N., Shouval, D., and Shafritz, D.A., 1980. "Identification of Integrated Hepatitis B Virus DNA and Expression of Viral RNA in an HBsAg-producing Human Hepatocellular Carcinoma Cell Line." *Nature*, **286**(5772): 531–3. 104

Chellappan, S., Kraus, V.B., Kroger, B., Munger, K., Howley, P.M., Phelps, W.C., and Nevins, J.R., 1992. "Adenovirus EtA, Simian Virus 40 Tumor Antigen, Human Papillomavirus E7 Protein Share the Capacity to Disrupt the Interaction Between Transcription Factor E2F and the Retinoblastoma Gene Product." *Proceedings of the National Academy of Sciences of the United States of America*, **89**(10): 4549–53. 73

Chen, T.M., Pecoraro, G., and Defendi, V., 1993. "Genetic Analysis of *in vitro* Progression of Human Papillomavirus-Transfected Human Cervical Cells." *Cancer Research*, **53**(5): 1167–71. 73

Chen, Y., 2001. "Glycogen Storage Disease." In C.R. Scriver, A.L. Beaudet, W.S. Sly, D. Valle, B. Vogelstein, and B. Childs, editors, *The Metabolic and Molecular Basis of Inherited Disease*, New York, NY: McGraw-Hill. 8th edition, 1521–51. 95, 96

des Chene, D., 2005. "Mechanisms of Life in the Seventeenth Century: Borelli, Perrault, Régis." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, **36**(2): 245–60. 110

Christiansen, G.W., 1938. "Lymphosarcoma of the Jaws and Palate." *Journal of the American Dental Association*, **25**(5): 728–30. 38

Churchill, A.E. and Biggs, P.M., 1967. "Agent of Marek's Disease in Tissue Culture." *Nature*, **215**(5100): 528–30. 16

Cleator, G. and Klapper, P., 2004a. "Herpes Simplex." In A.J. Zuckerman, J.E. Banatvala, J.R. Pattison, P. Griffiths, and B. Schoub, editors, *Principles and Practice of Clinical Virology*, Chichester: John Wiley & Sons. 5th edition, 27–51. 60

———, 2004b. "The Herpesviridae." In A.J. Zuckerman, J.E. Banatvala, J.R. Pattison, P. Griffiths, and B. Schoub, editors, *Principles and Practice of Clinical Virology*, Chichester: John Wiley & Sons. 5th edition, 23–6. 170

Clifford, P., 1961. "Malignant Disease of the Nose, Paranasal Sinuses and Post-Nasal Space in East Africa." *The Journal of Laryngology and Otology*, **75**(8): 707–33. 35

Clift, R.A., Wright, D.H., and Clifford, P., 1963. "Leukemia in Burkitt's Lymphoma." *Blood*, **22**(3): 243–51. 34

Coady, A., 1975. "Letter: The Aflatoxin-Hepatoma-HBAg Story." *British Medical Journal*, **3**(5983): 592–3. 103

Coggin, J. and zur Hausen, H., 1979. "Meeting Report: Workshop on Papillomaviruses and Cancer." *Cancer Research*, **39**(2): 545–6. 70, 72

Collard, W., Thornton, H., and Green, M., 1973. "Cells Transformed by Human Herpesvirus Type 2 Transcribe Virus-Specific RNA Sequences Shared by Herpesvirus Types 1 and 2." *Nature: New Biology*, **243**(130): 264–6. 62

Cori, G.T. and Green, A.A., 1943. "Crystalline Muscle Phosphorylase II. Prosthetic Group." *Journal of Biological Chemistry*, **151**(1): 31–8. 97

Cornfield, J., 1976. "Recent Methodological Contributions to Clinical Trials." *American Journal of Epidemiology*, **104**(4): 408–21. 19

Craver, C.F., 2001. "Role Functions, Mechanisms, and Hierarchy." *Philosophy of Science*, **68**(1): 53–74. 110, 111, 114

———, 2002. "Interlevel Experiments and Multilevel Mechanisms in the Neuroscience of Memory." *Philosophy of Science*, **69**(S3): 83–97. 110

———, 2005. "Beyond Reduction: Mechanisms, Multifield Integration and the Unity of Neuroscience." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, **36**(2): 373–95. 113, 137

———, 2007. *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford: Clarendon Press. 14, 110, 111, 112, 113, 137, 147, 161, 166

Craver, C.F. and Darden, L., 2005. "Introduction." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, **36**(2): 233–44. 110

Craver, L.F. and Copeland, M.M., 1934. "Lymphosarcoma in Bone." *Archives of Surgery*, **28**(5): 809–24. 38

Crawford, L. and Crawford, E., 1963. "A Comparative Study of Polyoma and Papilloma Viruses." *Virology*, **21**(2): 258–63. 68, 69

Crupi, V., Tentori, K., and Gonzalez, M., 2007. "On Bayesian Measures of Evidential Support: Theoretical and Empirical Issues." *Philosophy of Science*, **74**(2): 229–52. 19

Dalldorf, G., 1962. "Lymphomas of African Children with Different Forms or Environmental Influences." *Journal of the American Medical Association*, **181**(12): 1026–8. 36

Dalldorf, G. and Bergamini, F., 1964. "Unidentified, Filtrable Agents Isolated from African Children with Malignant Lymphomas." *Proceedings of the National Academy of Sciences of the United States of America*, **51**(2): 263–5. 48

Dalldorf, G., Bergamini, F., and Frost, P., 1966. "Further Observations of the Lymphomas of African Children." *Proceedings of the National Academy of Sciences of the United States of America*, **55**(2): 297–302. 48

Dane, D.S., Cameron, C.H., and Briggs, M., 1970. "Virus-Like Particles in Serum of Patients with Australia-Antigen-Associated Hepatitis." *The Lancet*, **295**(7649): 695–8. 102

Darden, L., 1982. "Artificial Intelligence and Philosophy of Science: Reasoning by Analogy in Theory Construction." *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, **1982**(2): 147–65. 166

———, 2002. "Strategies for Discovering Mechanisms: Schema Instantiation, Modular Subassembly, Forward/Backward Chaining." *Philosophy of Science*, **69**(S3): 354–65. 110

———, 2005. "Relations among Fields: Mendelian, Cytological and Molecular Mechanisms." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, **36**(2): 349–71. 110

———, 2006. *Reasoning in Biological Discoveries: Essays on Mechanisms, Interfield Relations, and Anomaly Resolution*. Cambridge: Cambridge University Press. 110, 153, 161, 166

Dargeon, H.W., 1961. "Lymphosarcoma in Childhood." *American Journal of Roentgenology, Radium Therapy and Nuclear Medicine*, **85**(4): 729–32. 38

Davies, J.N., Elmes, S., Hutt, M.S., Mtimavalye, L.A., Owor, R., and Shaper, L., 1964. "Cancer in an African Community, 1897–1956. An Analysis of the Records of Mengo Hospital, Kampala, Uganda. Part 2." *British Medical Journal*, **1**(5379): 336–41. 31

Davies, J.N., Wilson, B.A., and Knowelden, J., 1958. "Cancer in Kampala: A Survey in an Underdeveloped Country." *British Medical Journal*, **2**(5093): 439–43. 35

Dawid, A.P., 2002. "Counterfactuals: Help or Hindrance?" *International Journal of Epidemiology*, **31**(2): 429–30; discussion 435–8. 19

de-Thé, G., Geser, A., Day, N.E., Tukei, P.M., Williams, E.H., Beri, D.P., Smith, P.G., Dean, A.G., Bronkamm, G.W., Feorino, P., and Henle, W., 1978. "Epidemiological Evidence for Causal Relationship Between Epstein-Barr Virus and Burkitt's Lymphoma from Ugandan Prospective Study." *Nature*, **274**(5673): 756–61. 31, 51, 94, 95, 170, 180

Deaver, J.B. and Reimann, S.P., 1931. "Cancer of the Uterus: with General Remarks on the Pathologic Aspects of Cancer of the Uterus." *Annals of Surgery*, **94**(3): 381–8. 56

Deinhardt, F., 1973. "Symposium on Herpesvirus and Cervical Cancer. Epilog." *Cancer Research*, **33**(6): 1556. 62, 63

Delap, R.D., Friedman-Kien, A., and Rush, M.G., 1976. "The Absence of Human Papilloma Viral DNA Sequences in Condylomata Acuminata." *Virology*, **74**(1): 268–72. 69

Delescluse, C., Prunieras, M., Regnier, M., Moreno, G., and Arouete, J., 1972. "Epidermodysplasia Verruciformis. I. Electron Microscope Autoradiography and Tissue Culture Studies." *Archiv Für Dermatologische Forschung*, **242**(2): 202–15. 67

Deyrup, A.T., 2008. "Epstein-Barr Virus-Associated Epithelial and Mesenchymal Neoplasms." *Human Pathology*, **39**(4): 473–83. 28

DiPaolo, J.A., Woodworth, C.D., Popescu, N.C., Koval, D.L., Lopez, J.V., and Doniger, J., 1990. "HSV-2-Induced Tumorigenicity in HPV16-Immortalized Human Genital Keratinocytes." *Virology*, **177**(2): 777–9. 65

Doll, R. and Peto, R., 1981. *The Causes of Cancer: Quantitative Estimates of Avoidable Risks of Cancer in America Today*. Oxford: Oxford University Press. 19

Dorn, H., 1953. "Philosophy of Inferences from Retrospective Studies." *American Journal of Public Health and the Nation's Health*, **43**(1): 677–83. 19

Dowe, P., 1992. "Wesley Salmon's Process Theory of Causality and the Conserved Quantity Theory." *Philosophy of Science*, **59**(2): 195–216. 134, 135

———, 1995. "Causality and Conserved Quantities: A Reply to Salmon." *Philosophy of Science*, **62**(2): 321–33. 135

———, 2000. *Physical Causation*. Cambridge: Cambridge University Press. 82, 110, 134

———, 2001. "A Counterfactual Theory of Prevention and 'Causation' by Omission." *Australasian Journal of Philosophy*, **79**(2): 216–26. 146

Down, L., 1866. "Observations on an Ethnic Classification of Idiots." *London Hospital Reports*, **3**: 259–62. 155

Dudley, F.J., Giustino, V., and Sherlock, S., 1972. "Cell-mediated Immunity in Patients Positive for Hepatitis-associated Antigen." *British Medical Journal*, **4**(5843): 754–6. 103

Duff, R. and Rapp, F., 1971. "Properties of Hamster Embryo Fibroblasts Transformed *in vitro* after Exposure to Ultraviolet-Irradiated Herpes Simplex Virus Type 2." *Journal of Virology*, **8**(4): 469–77. 62, 63

———, 1973. "Oncogenic Transformation of Hamster Embryo Cells After Exposure to Inactivated Herpes Simplex Virus Type 1." *Journal of Virology*, **12**(2): 209–17. 62

———, 1975. "Quantitative Assay for Transformation of 3T3 Cells by Herpes Simplex Virus Type 2." *Journal of Virology*, **15**(3): 490–6. 62

Dunn, A.E.G. and Ogilvie, M.M., 1968. "Intranuclear Virus Particles in Human Genital Wart Tissue: Observations on the Ultrastructure of the Epidermal Layer." *Journal of Ultrastructure Research*, **22**(3/4): 282–95. 67

Dürst, M., Dzarlieva-Petrusevska, R.T., Boukamp, P., Fusenig, N.E., and Gissmann, L., 1987. "Molecular and Cytogenetic Analysis of Immortalized Human Primary Keratinocytes Obtained After Transfection with Human Papillomavirus Type 16 DNA." *Oncogene*, **1**(3): 251–6. 73

Dürst, M., Gissmann, L., Ikenberg, H., and zur Hausen, H., 1983. "A Papillomavirus DNA from a Cervical Carcinoma and its Prevalence in Cancer Biopsy Samples from Different Geographic Regions." *Proceedings of the National Academy of Sciences of the United States of America*, **80**(12): 3812–5. 71, 74

Dyson, N., Howley, P.M., Münger, K., and Harlow, E., 1989. "The Human Papilloma Virus-16 E7 Oncoprotein is able to Bind to the Retinoblastoma Gene Product." *Science*, **243**(4893): 934–7. 73

Eddy, B.E., 1962. "Tumors Produced in Hamsters by SV40." *Federation Proceedings*, **21**: 930–5. 17

Eells, E., 1991. *Probabilistic Causality*. Cambridge: Cambridge University Press. 108

Egan, R.L. and Dodd, G.D., 1957. "Roentgen Characteristics of the Malignant Lymphomas and Leukemias in Childhood." *Texas State Journal of Medicine*, **53**(10): 775–81. 38

Einfeld, S.L. and Brown, R., 2010. "Down Syndrome–New Prospects for an Ancient Disorder." *Journal of the American Medical Association*, **303**(24): 2525–6. 155

Ellermann, V. and Bang, O., 1909. "Experimentelle Leukämie bei Hühnern [Experimental Leukemia in Chickens]." *Zeitschrift für Hygiene und Infektionskrankheiten*, **63**(1): 231–72. 16

Elliott, R., 1964. "On the Prevention of Carcinoma of the Cervix." *The Lancet*, **1**(7327): 231–5. 57, 58, 59

Elwood, J.M., 1988. *Causal Relationships in Medicine*. Oxford: Oxford University Press. 19, 166

Epstein, M.A., 1956. "Intra-Cellular Identification of the Rous Virus." *Nature*, **178**(4523): 45–6. 42

———, 1958a. "Composition of the Rous Virus Nucleoid." *Nature*, **181**(4626): 1808. 42

———, 1958b. "Observations on the Rous Virus; Purification and Identification of the Particles from Solid Tumours." *British Journal of Cancer*, **12**(2): 248–55. 42

———, 1962a. "Observations on the Fine Structure of Mature Herpes Simplex Virus and on the Composition of Its Nucleoid." *The Journal of Experimental Medicine*, **115**(1): 1–12. 42, 45

———, 1962b. "Observations on the Mode of Release of Herpes Virus from Infected HeLa Cells." *The Journal of Cell Biology*, **12**(3): 589–97. 42

———, 2001a. "Historical Background." *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*, **356**(1408): 413–20. 15, 26

———, 2001b. "Reflections on Epstein-Barr Virus: Some Recently Resolved Old Uncertainties." *Journal of Infection*, **43**(2): 111–5. 51

———, 2005. "The Origins of EBV Research: Discovery and Characterization of the Virus." In E.S. Robertson, editor, *Epstein-Barr Virus*, Wymondham, Norfolk: Caister Academic Press. 1–14. 26, 42, 43, 44, 45, 49

Epstein, M.A., Achong, B.G., and Barr, Y.M., 1964a. "Virus Particles in Cultured Lymphoblasts from Burkitt's Lymphoma." *The Lancet*, **283**(7335): 702–3. 44, 45

Epstein, M.A., Achong, B.G., Barr, Y.M., Zajac, B., Henle, G., and Henle, W., 1966. "Morphological and Virological Investigations on Cultured Burkitt Tumor Lymphoblasts (Strain Raji)." *Journal of the National Cancer Institute*, **37**(4): 547–59. 45

Epstein, M.A., Achong, B.G., Churchill, A.E., and Biggs, P.M., 1968. "Structure and Development of the Herpes-type Virus of Marek's Disease." *Journal of the National Cancer Institute*, **41**(3): 805–20. 16

Epstein, M.A. and Barr, Y.M., 1964. "Cultivation *in vitro* of Human Lymphoblasts from Burkitt's Malignant Lymphoma." *The Lancet*, **283**(7327): 252–3. 43, 44, 46

Epstein, M.A., Barr, Y.M., and Achong, B.G., 1964b. "A Second Virus-Carrying Tissue Culture Strain (EB2) of Lymphoblasts from Burkitt's Lymphoma." *Pathologie-Biologie*, **12**(23/4): 1233–4. 45, 46

———, 1965a. "Studies with Burkitt's Lymphoma." *The Wistar Institute Symposium Monograph*, **4**: 69–82. 44, 45, 46, 162

Epstein, M.A. and Eastwood, M.A., 1995. "Denis Parsons Burkitt. 28 February 1911–23 March 1993." *Biographical Memoirs of Fellows of the Royal Society*, **41**: 88–102. 29, 30

Epstein, M.A., Henle, G., Achong, B.G., and Barr, Y.M., 1965b. "Morphological and Biological Studies on a Virus in Cultured Lymphoblasts from Burkitt's Lymphoma." *Journal of Experimental Medicine*, **121**(5): 761–70. 47, 48, 49

Epstein, M.A. and Herdson, P.B., 1963. "Cellular Degeneration Associated with Characteristic Nuclear Fine Structural Changes in the Cells from Two Cases of Burkitt's Lymphoma Syndrome." *British Journal of Cancer*, **17**(1): 56–8. 43

Epstein, M.A., Hummeler, K., and Berkaloff, A., 1964c. "The Entry and Distribution of Herpes Virus and Colloidal Gold in Hela Cells After Contact in Suspension." *The Journal of Experimental Medicine*, **119**(2): 291–302. 42

Epstein, M., Barr, Y., and Achong, B., 1964d. "Avian Tumor Virus Behavior as a Guide in the Investigation of a Human Neoplasm." In J.W. Beard, editor, *International Conference on Avian Tumor Viruses*, Bethesda, MA: National Institutes of Health, volume 17. 637–50. 43

Evans, A.S., 1976. "Causation and Disease: The Henle-Koch Postulates Revisited." *The Yale Journal of Biology and Medicine*, **49**(2): 175–95. 17, 18, 19, 166

———, 1977. "Limitation of Koch's Postulates." *The Lancet*, **301**(8051): 1277–8. 18

———, 1978. "Causation and Disease: A Chronological Journey. The Thomas Parran Lecture." *American Journal of Epidemiology*, **108**(4): 249–58. 18, 19

———, 1993. *Causation and Disease: A Chronological Journey*. New York, NY: Plenum Press. 17

Evans, A.S. and Niederman, J.C., 1989. "Epstein-Barr Virus." In A.S. Evans and R.A. Kaslow, editors, *Viral Infections of Humans*, New York, NY: Plenum Press. 265–92. 27

Evans, D., 2003. "Hierarchy of Evidence: A Framework for Ranking Evidence Evaluating Healthcare Interventions." *Journal of Clinical Nursing*, **12**(1): 77–84. 84

Fair, D., 1979. "Causation and the Flow of Energy." *Erkenntnis*, **14**(3): 219–50. 146

Falkow, S., 1988. "Molecular Koch's Postulates Applied to Microbial Pathogenicity." *Reviews of Infectious Diseases*, **10**(Suppl 2): S274–6. 18

Favre, M., Orth, G., Croissant, O., and Yaniv, M., 1975. "Human Papillomavirus DNA: Physical Map." *Proceedings of the National Academy of Sciences of the United States of America*, **72**(12): 4810–4. 69, 72

Fawcett, D., 1956. "Electron Microscope Observations on Intracellular Virus-Like Particles Associated with the Cells of the Lucké Renal Adenocarcinoma." *The Journal of Biophysical and Biochemical Cytology*, **2**(6): 725–41. 16

Felsher, Z., 1947. "Virus Studies on Cutaneous Material." *The Journal of Investigative Dermatology*, **8**(3): 123–4. 67

Findlay, G., 1930. "Warts." In *A System of Bacteriology in Relation to Medicine*, London: H.M. Stationery Office, volume 7, chapter XVIII. 252–8. 66

Findlay, G. and MacCallum, F., 1937. "Note on Acute Hepatitis and Yellow Fever Immunization." *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **31**(3): 297–308. 101

Fink, M.A., Malmgren, R.A., Rauscher, F.J., Orr, H.C., and Karon, M., 1964. "Application of Immunofluorescence to the Study of Human Leukemia." *Journal of the National Cancer Institute*, **33**: 581–8. 46

Fischer, E., 1953. "Versuche zur Kultivierung des 'Warzenvirus' auf der Chorioallantois des Huhnereies [Attempts to Cultivate the 'Wart Virus' on the Chorioallantois of Hens' Eggs]." *Archiv für Klinische und Experimentelle Dermatologie*, **196**(3): 256–60. 67

Fischer, G.A., 1957. "Tissue Culture of Mouse Leukemic Cells." *Proceedings of the American Association for Cancer Research*, **2**: 201. 43

———, 1958. "Studies of the Culture of Leukemic Cells *in vitro*." *Annals of the New York Academy of Sciences*, **76**(Screening Procedures for Experimental Cancer Chemotherapy): 673–80. 43

Fish, E.N., Tobin, S.M., Cooter, N.B., and Papsin, F.R., 1982. "Update on the Relation of Herpesvirus Hominis type II to Carcinoma of the Cervix." *Obstetrics and Gynecology*, **59**(2): 220–4. 62

Flanagan, N., Healy, E., Ray, A., Philips, S., Todd, C., Jackson, I.J., Birch-Machin, M.A., and Rees, J.L., 2000. "Pleiotropic Effects of the Melanocortin 1 Receptor (MC1R) Gene on Human Pigmentation." *Human Molecular Genetics*, **9**(17): 2531–7. 131

Fouchier, R.A.M., Kuiken, T., Schutten, M., van Amerongen, G., van Doornum, G.J.J., van den Hoogen, B.G., Peiris, M., Lim, W., Stöhr, K., and Osterhaus, A.D.M.E., 2003. "Aetiology: Koch's Postulates Fulfilled for SARS Virus." *Nature*, **423**(6937): 240. 18

Francis, D.P., Hadler, S.C., Thompson, S.E., Maynard, J.E., Ostrow, D.G., Altman, N., Braff, E.H., O'Malley, P., Hawkins, D., Judson, F.N., Penley, K., Nylund, T., Christie, G., Meyers, F., Moore, J.N., Gardner, A., Doto, I.L., Miller, J.H., Reynolds, G.H., Murphy, B.L., Schable, C.A., Clark, B.T., Curran, J.W., and Redeker, A.G., 1982. "The Prevention of Hepatitis B with Vaccine. Report of the Centers for Disease Control Multi-Center Efficacy Trial Among Homosexual Men." *Annals of Internal Medicine*, **97**(3): 362–6. 104

Fraumeni, J.F., Lloyd, J.W., Smith, E.M., and Wagoner, J.K., 1969. "Cancer Mortality Among Nuns: Role of Marital Status in Etiology of Neoplastic Disease in Women." *Journal of the National Cancer Institute*, **42**(3): 455–68. 56

Fredericks, D.N. and Relman, D.A., 1996. "Sequence-based Identification of Microbial Pathogens: A Reconsideration of Koch's Postulates." *Clinical Microbiology Reviews*, **9**(1): 18–33. 18

Frenkel, N., Locker, H., Cox, B., Roizman, B., and Rapp, F., 1976. "Herpes Simplex Virus DNA in Transformed Cells: Sequence Complexity in Five Hamster Cell Lines and One Derived Hamster Tumor." *Journal of Virology*, **18**(3): 885–93. 62

Frenkel, N., Roizman, B., Cassai, E., and Nahmias, A., 1972. "A DNA Fragment of Herpes Simplex 2 and Its Transcription in Human Cervical Cancer Tissue." *Proceedings of the National Academy of Sciences of the United States of America*, **69**(12): 3784–9. 61, 62

Friedberg, M. and Serlin, O., 1963. "Condyloma Acuminatum." *Diseases of the Colon & Rectum*, **6**(5): 352–5. 70

Gagnon, F., 1950. "Contribution to the Study of the Etiology and Prevention of Cancer of the Cervix of the Uterus." *American Journal of Obstetrics and Gynecology*, **60**(3): 516–22. 56

Galloway, D.A. and McDougall, J.K., 1983. "The Oncogenic Potential of Herpes Simplex Viruses: Evidence for a 'Hit-and-Run' Mechanism." *Nature*, **302**(5903): 21–4. 65

Galloway, J., 1989. "Epstein-Barr Virus. Chance and Felicity." *Nature*, **338**(6215): 463–4. 26

Gardner, E.J., 1951. "A Genetic and Clinical Study of Intestinal Polyposis, a Predisposing Factor for Carcinoma of the Colon and Rectum." *American Journal of Human Genetics*, **3**(2): 167–76. 159

Gardner, J.W. and Lyon, J.L., 1977. "Low Incidence of Cervical Cancer in Utah." *Gynecologic Oncology*, **5**(1): 68–80. 56

Gerber, P., Lucas, S., Nonoyama, M., Perlin, E., and Goldstein, L.I., 1972. "Oral Excretion of Epstein-Barr Virus by Healthy Subjects and Patients with Infectious Mononucleosis." *The Lancet*, **300**(7785): 988–9. 27

Gillies, D., 2000. *Philosophical Theories of Probability*. Abingdon: Routledge. 174

———, 2007. "Review: K. Codell Carter: The Rise of Causal Concepts of Disease. Case Histories." *British Journal for the Philosophy of Science*, **58**(2): 365–77. 18

———, 2010. "The Russo-Williamson Thesis and the Question of whether Smoking Causes Heart Disease." In P. McKay Illari, F. Russo, and J. Williamson, editors, *Causality in the Sciences*, Oxford: Oxford University Press. Forthcoming. 86

Girardi, A.J., Hilleman, M.R., and Zwickey, R.E., 1964. "Tests in Hamsters for Oncogenic Quality of Ordinary Viruses Including Adenovirus Type 7." *Proceedings of the Society for Experimental Biology and Medicine*, **115**(4): 1141–50. 17

Gissmann, L., Diehl, V., Schultz-Coulon, H.J., and zur Hausen, H., 1982a. "Molecular Cloning and Characterization of Human Papilloma Virus DNA Derived from a Laryngeal Papilloma." *Journal of Virology*, **44**(1): 393–400. 71

Gissmann, L. and zur Hausen, H., 1976. "Human Papilloma Virus DNA: Physical Mapping and Genetic Heterogeneity." *Proceedings of the National Academy of Sciences of the United States of America*, **73**(4): 1310–3. 69

———, 1980. "Partial Characterization of Viral DNA from Human Genital Warts (Condylomata Acuminata)." *International Journal of Cancer*, **25**(5): 605–9. 70, 71

Gissmann, L., Pfister, H., and zur Hausen, H., 1977. "Human Papilloma Viruses (HPV): Characterization of Four Different Isolates." *Virology*, **76**(2): 569–80. 70, 72

Gissmann, L., de Villiers, E.M., and zur Hausen, H., 1982b. "Analysis of Human Genital Warts (Condylomata Acuminata) and Other Genital Tumors for Human Papillomavirus Type 6 DNA." *International Journal of Cancer*, **29**(2): 143–6. 71

Gissmann, L., Wolnik, L., Ikenberg, H., Koldovsky, U., Schnrch, H.G., and zur Hausen, H., 1983. "Human Papillomavirus Types 6 and 11 DNA Sequences in Genital and Laryngeal Papillomas and in Some Cervical Cancers." *Proceedings of the National Academy of Sciences of the United States of America*, **80**(2): 560–3. 71

Glaser, R., Padgett, D.A., Litsky, M.L., Baiocchi, R.A., Yang, E.V., Chen, M., Yeh, P., Klimas, N.G., Marshall, G.D., Whiteside, T., Herberman, R., Kiecolt-Glaser, J., and Williams, M.V., 2005. "Stress-Associated Changes in the Steady-State Expression of Latent Epstein-Barr Virus: Implications for Chronic Fatigue Syndrome and Cancer." *Brain, Behavior, and Immunity*, **19**(2): 91–103. 28

Glemser, B., 1971. *The Long Safari*. London: Bodley Head. 28, 29, 30, 31, 32, 34

Glennan, S.S., 1996. "Mechanisms and the Nature of Causation." *Erkenntnis*, **44**(1): 49–71. 115, 116, 117, 129

———, 2002. "Rethinking Mechanistic Explanation." *Philosophy of Science*, **69**(S3): 342–53. 115, 121, 132

———, 2005. "Modeling Mechanisms." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, **36**(2): 443–64. 121, 122, 132

Granger, C.W.J., 1969. "Investigating Causal Relations by Econometric Models and Cross-spectral Methods." *Econometrica*, **37**(3): 424–38. 108

Gravell, M., 1971. "Viruses and Renal Carcinoma of *Rana pipiens*: X. Comparison of Herpes-type Viruses Associated with Lucké Tumor-bearing Frogs." *Virology*, **43**(3): 730–3. 17

Gravitt, P.E., Schiffman, M., Solomon, D., Wheeler, C.M., and Castle, P.E., 2008. "A Comparison of Linear Array and Hybrid Capture 2 for Detection of Carcinogenic Human Papillomavirus and Cervical Precancer in ASCUC-LSIL Triage Study." *Cancer Epidemiology, Biomarkers & Prevention*, **17**(5): 1248–54. 75

Green, A.A., Cori, G.T., and Oncley, J.L., 1943. "Crystalline Muscle Phosphorylase. I. Preparation, Properties, and Molecular Weight." *Journal of Biological Chemistry*, **151**(1): 21–9. 97

Green, M., Brackmann, K.H., Sanders, P.R., Loewenstein, P.M., Freel, J.H., Eisinger, M., and Switlyk, S.A., 1982. "Isolation of a Human Papillomavirus from a Patient with Epidermodysplasia Verruciformis: Presence of Related Viral DNA Genomes in Human Urogenital Tumors." *Proceedings of the National Academy of Sciences of the United States of America*, **79**(14): 4437–41. 71

Greenland, S. and Brumback, B., 2002. "An Overview of Relations Among Causal Modelling Methods." *International Journal of Epidemiology*, **31**(5): 1030–7. 19

Greenland, S. and Poole, C., 1988. "Invariants and Noninvariants in the Concept of Interdependent Effects." *Scandinavian Journal of Work, Environment & Health*, **14**(2): 125–9. 19

Griffin, E.R., Wright, D.H., Bell, T.M., and Ross, M.G., 1966. "Demonstration of Virus Particles in Biopsy Material from Cases of Burkitt's Tumour." *European Journal of Cancer*, **2**(4): 353–8. 48

Grimes, D., 2006. "Kochs Postulates–Then and Now." *Microbe*, **1**(5): 223. 18

Gross, L., 1952. "Mouse Leukemia." *Annals of the New York Academy of Sciences*, **54**(Viruses as Causative Agents in Cancer): 1184–96. 17

Gross, L., McCarty, K.S., and Gessler, A.E., 1952. "The Significance of Particles in Human Milk." *Annals of the New York Academy of Sciences*, **54**(Viruses as Causative Agents in Cancer): 1018–34. 16

Guyatt, G.H., Oxman, A.D., Vist, G.E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., and Schnemann, H.J., 2008. "GRADE: An Emerging Consensus on Rating Quality of Evidence and Strength of Recommendations." *BMJ*, **336**(7650): 924–6. 84

Hackney, J.D. and Linn, W.S., 1979. "Koch's Postulates Updated: A Potentially Useful Application to Laboratory Research and Policy Analysis in Environmental Toxicology." *The American Review of Respiratory Disease*, **119**(6): 849–52. 18

Haddow, A.J., 1964. "Age Incidence in Burkitt's Lymphoma Syndrome." *East African Medical Journal*, **41**(1): 1–6. 48

Haddow, A.J. and McCallum, D., 1962. "Preliminary Statistical Work on Mr. Burkitt's Records." In A.J. Haddow and D. McCallum, editors, *East African Virus Research Institute Report 1961–2*, Nairobi Government Printer. 31–6. 47

Hadziyannis, S.J., Merikas, G.E., and Afroudakis, A.P., 1970. "Hepatitis-associated Antigen in Chronic Liver Disease." *The Lancet*, **1**(7663): 100. 102

Hájek, A., 2007. "The Reference Class Problem Is Your Problem Too." *Synthese*, **156**(3): 563–85. 175

Halbert, C.L., Demers, G.W., and Galloway, D.A., 1991. "The E7 Gene of Human Papillomavirus Type 16 Is Sufficient for Immortalization of Human Epithelial Cells." *Journal of Virology*, **65**(1): 473–8. 73

Hall, P.A. and Lemoine, N.R., 1991. "Koch's Postulates Revisited." *The Journal of Pathology*, **164**(4): 283–4. 18

Hamilton, M.A., 1979. "Choosing the Parameter for a 2 x 2 Table or a 2 x 2 x 2 Table Analysis." *American Journal of Epidemiology*, **109**(3): 362–75. 19

Hammond, E., 1955. "Cause and Effect." In E.S. Wynder, editor, *The Biologic Effect of Tobacco*, London: J.A. Churchill, chapter 7. 171–96. 19

Hampar, B. and Ellison, S.A., 1963. "Cellular Alterations in the MCH Line of Chinese Hamster Cells Following Infection with Herpes Simplex Virus." *Proceedings of the National Academy of Sciences of the United States of America*, **49**(4): 474–80. 61

Harding, W.G., 1942. "The Influence of Syphilis in Cancer of the Cervix Uteri." *Cancer Research*, **2**(1): 59–61. 58

Harris, N.L., Jaffe, E.S., Diebold, J., Flandrin, G., Muller-Hermelink, H.K., Vardiman, J., Lister, T.A., and Bloomfield, C.D., 1999. "World Health Organization Classification of Neoplastic Diseases of the Hematopoietic and Lymphoid Tissues: Report of the Clinical Advisory Committee Meeting - Airlie House, Virginia, November 1997." *Journal of Clinical Oncology*, **17**(12): 3835–49. 26

zur Hausen, H., 1975. "Oncogenic Herpes viruses." *Biochimica Et Biophysica Acta*, **417**(1): 25–53. 17, 65

———, 1976. "Condylomata Acuminata and Human Genital Cancer." *Cancer Research*, **36**(2 pt 2): 794. 69

———, 1980. "The Role of Viruses in Human Tumors." *Advances in Cancer Research*, **33**: 77–107. 15

———, 1982. "Human Genital Cancer: Synergism Between Two Virus Infections or Synergism Between a Virus Infection and Initiating Events?" *The Lancet*, **2**(8312): 1370–2. 65

———, 1983. "Herpes Simplex Virus in Human Genital Cancer." *International Review of Experimental Pathology*, **25**: 307–26. 69, 71

———, 1986. "Intracellular Surveillance of Persisting Viral Infections. Human Genital Cancer Results from Deficient Cellular Control of Papillomavirus Gene Expression." *The Lancet*, **2**(8505): 489–91. 73

———, 1989a. "Papillomavirus in Anogenital Cancer: The Dilemma of Epidemiologic Approaches." *Journal of the National Cancer Institute*, **81**(22): 1680–2. 74

———, 1989b. "Papillomaviruses in Anogenital Cancer as a Model to Understand the Role of Viruses in Human Cancers." *Cancer Research*, **49**(17): 4677–81. 54, 74

———, 1991. "Human Papillomaviruses in the Pathogenesis of Anogenital Cancer." *Virology*, **184**(1): 9–13. 73

———, 1996a. "Papillomavirus Infections–a Major Cause of Human Cancers." *Biochimica Et Biophysica Acta*, **1288**(2): F55–78. 66, 73

———, 1996b. "Roots and Perspectives of Contemporary Papillomavirus Research." *Journal of Cancer Research and Clinical Oncology*, **122**(1): 3–13. 66, 73

———, 2000. "Papillomaviruses Causing Cancer: Evasion from Host-Cell Control in Early Events in Carcinogenesis." *Journal of the National Cancer Institute*, **92**(9): 690–8. 73

———, 2006. *Infections Causing Human Cancer*. Weinheim: Wiley–VCH. 15, 26, 101, 104

———, 2008. "Harald zur Hausen." URL `http://nobelprize.org/nobel_prizes/medicine/laureates/2008/hausen.html`. 70

zur Hausen, H., Meinhof, W., Scheiber, W., and Bornkamm, G., 1974a. "Attempts to Detect Virus-Specific DNA in Human Tumors. I. Nucleic Acid Hybridizations with Complementary RNA of Human Wart Virus." *International Journal of Cancer*, **13**(5): 650–6. 69

zur Hausen, H. and Schulte-Holthausen, H., 1970. "Presence of EB Virus Nucleic Acid Homology in a 'Virus-Free' Line of Burkitt Tumour Cells." *Nature*, **227**(5255): 245–8. 50

zur Hausen, H., Schulte-Holthausen, H., Klein, G., Henle, W., Henle, G., Clifford, P., and Santesson, L., 1970. "EBV DNA in Biopsies of Burkitt Tumours and Anaplastic Carcinomas of the Nasopharynx." *Nature*, **228**(5276): 1056–8. 28, 49, 50

zur Hausen, H., Schulte-Holthausen, H., Wolf, H., Drries, K., and Egger, H., 1974b. "Attempts to Detect Virus-Specific DNA in Human Tumors. II. Nucleic Acid Hybridizations with Complementary RNA of Human Herpes Group Viruses." *International Journal of Cancer*, **13**(5): 657–64. 64

zur Hausen, H. and de Villiers, E.M., 1994. "Human Papillomaviruses." *Annual Review of Microbiology*, **48**: 427–47. 54

Hausman, D. and Woodward, J., 1999. "Independence, Invariance and the Causal Markov Condition." *British Journal for the Philosophy of Science*, **50**(4): 521–83. 131

Heaton, K., 1993. "Obituary: Denis Burkitt." *The Lancet*, **341**(8850): 951–2. 29, 38

Hecht, J.L. and Aster, J.C., 2000. "Molecular Biology of Burkitt's Lymphoma." *Journal of Clinical Oncology*, **18**(21): 3707–21. 27

Hellwig, C.A., 1947. "Malignant Lymphoma: The Value of Radical Surgery in Selected Cases." *Surgery, Gynaecology and Obstetrics*, **84**(5): 950–8. 38

Henle, G. and Henle, W., 1966a. "Studies on Cell Lines Derived from Burkitt's Lymphoma." *Transactions of the New York Academy of Sciences*, **29**(1): 71–9. 49

Henle, G., Henle, W., Clifford, P., Diehl, V., Kafuko, G.W., Kirya, B.G., Klein, G., Morrow, R.H., Munube, G.M., Pike, P., Tukei, P.M., and Ziegler, J.L., 1969. "Antibodies to Epstein-Barr Virus in Burkitt's Lymphoma and Control Groups." *Journal of the National Cancer Institute*, **43**(5): 1147–57. 27, 51

Henle, G., Henle, W., and Diehl, V., 1968. "Relation of Burkitt's Tumor-Associated Herpes-Type Virus to Infectious Mononucleosis." *Proceedings of the National Academy of Sciences of the United States of America*, **59**(1): 94–101. 49, 50

Henle, G. and Henle, W., 1966b. "Immunofluorescence in Cells Derived from Burkitt's Lymphoma." *Journal of Bacteriology*, **91**(3): 1248–56. 49

Henle, J., 1938. "On Miasmata and Contagia." *Bulletin of the Institute of Story of Medicine*, **6**(8): 911–83. 18

Henle, W., Diehl, V., Kohn, G., zur Hausen, H., and Henle, G., 1967. "Herpes-Type Virus and Chromosome Marker in Normal Leukocytes After Growth with Irradiated Burkitt Cells." *Science*, **157**(792): 1064–5. 49

Henle, W., Henle, G., Gunvn, P., Klein, G., Clifford, P., and Singh, S., 1973. "Patterns of Antibodies to Epstein-Barr Virus-Induced Early Antigens in Burkitt's Lymphoma. Comparison of Dying Patients with Long-Term Survivors." *Journal of the National Cancer Institute*, **50**(5): 1163–73. 64

Henle, W., Henle, G., Ho, H.C., Burtin, P., Cachin, Y., Clifford, P., de Schryver, A., de-Thé, G., Diehl, V., and Klein, G., 1970. "Antibodies to Epstein-Barr Virus in Nasopharyngeal Carcinoma, Other Head and Neck Neoplasms, and Control Groups." *Journal of the National Cancer Institute*, **44**(1): 225–31. 49

Herath, C.H.P. and Chetty, R., 2008. "Epstein-Barr Virus-associated Lymphoepithelioma-like Gastric Carcinoma." *Archives of Pathology & Laboratory Medicine*, **132**(4): 706–9. 28

Hesse, M.B., 1963. *Models and Analogies in Science*. London: Sheed & Ward. 166

Higenbottam, T., Shipley, M.J., and Rose, G., 1982. "Cigarettes, Lung Cancer, and Coronary Heart Disease: The Effects of Inhalation and Tar Yield." *Journal of Epidemiology and Community Health*, **36**(2): 113–7. 104

Higginson, J. and Oettlé, A.G., 1960. "Cancer Incidence in the Bantu and 'Cape Colored' Races of South Africa: Report of a Cancer Survey in the Transvaal (1953–55)." *Journal of the National Cancer Institute*, **24**(3): 589–671. 35, 36

Hill, A.B., 1965. "The Environment and Disease: Association or Causation?" *Proceedings of the Royal Society of Medicine*, **58**(5): 295–300. 19, 93, 166

Hilton-Jones, D., 2001. "McArdle's Disease." *Practical Neurology*, **1**(2): 122–5. 95

Hitchcock, C.R., 1995. "Salmon on Explanatory Relevance." *Philosophy of Science*, **62**(2): 304–20. 135

———, 2001. "A Tale of Two Effects." *Philosophical Review*, **110**(3): 361–6. 171

Hopwood, P. and Crawford, D.H., 2000. "The role of EBV in Post-transplant Malignancies: a Review." *Journal of Clinical Pathology*, **53**(4): 248–54. 28

Howick, J., 2010. "Exposing the Vanities – and a Qualified Defence – of Mechanistic Reasoning in Clinical Decision-Making." Unpublished MS draft. 86, 88, 89, 90, 91

Howley, P.M., 1991. "Role of the Human Papillomaviruses in Human Cancer." *Cancer Research*, **51**(18 Suppl): 5019s–22s. 54

Huebner, R.J., 1957. "Criteria for Etiologic Association of Prevalent Viruses with Prevalent Diseases; The Virologist's Dilemma." *Annals of the New York Academy of Sciences*, **67**(8): 430–8. 18

Huebner, R.J., Rowe, W.P., and Lane, W.T., 1962. "Oncogenic Effects in Hamsters of Human Adenovirus Types 12 and 18." *Proceedings of the National Academy of Sciences of the United States of America*, **48**(12): 2051–8. 17

Hume, D., 1975 (1748). *An Enquiry Concerning Human Understanding*. Oxford: Oxford University Press, 3rd edition. 108, 113

Hung, C., Ma, B., Monie, A., Tsen, S., and Wu, T., 2008. "Therapeutic Human Papillomavirus Vaccines: Current Clinical Trials and Future Directions." *Expert Opinion on Biological Therapy*, **8**(4): 421–439. 75

Hurlin, P.J., Kaur, P., Smith, P.P., Perez-Reyes, N., Blanton, R.A., and McDougall, J.K., 1991. "Progression of Human Papillomavirus Type 18-Immortalized Human Keratinocytes to a Malignant Phenotype." *Proceedings of the National Academy of Sciences of the United States of America*, **88**(2): 570–4. 73

Hutchison, R., 1907. "On Suprarenal Sarcoma in Children with Metastases in the Skull." *Quarterly Journal of Medicine*, **1**(1): 33–8. 35

Hutt, M.S., 1981. "Historical Introduction, Burkitt's Lymphoma, Nasopharyngeal Carcinoma and Kaposi's Sarcoma." *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **75**(6): 761–5. 31, 52

IARC Working Group on the Evaluation of Cancer Preventive Strategies, 2005. *Cervix Cancer Screening*, volume 10 of *IARC Handbooks of Cancer Prevention*. Lyon: IARCPress. 54

IARC Working Group on the Evaluation of Carcinogenic Risks to Humans, 1995. *Human Papillomaviruses*, volume 64 of *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans*. Lyon: IARCPress. 54

———, 2006. *Human Papillomaviruses*, volume 90 of *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans*. Lyon: IARCPress. 54

Inglis, T.J.J., 2007. "Principia Aetiologica: taking Causality beyond Koch's Postulates." *Journal of Medical Microbiology*, **56**(11): 1419–22. 18

Ito, Y. and Evans, C.A., 1961. "Induction of Tumors in Domestic Rabbits with Nucleic Acid Preparations from Partially Purified Shope Papilloma Virus and from Extracts of the Papillomas of Domestic and Cottontail Rabbits." *Journal of Experimental Medicine*, **114**(4): 485–500. 16

Iversen, O.H., Iversen, U., Ziegler, J.L., and Bluming, A.Z., 1974. "Cell Kinetics in Burkitt Lymphoma." *European Journal of Cancer*, **10**(3): 155–63. 27

Jablonska, S., Dabrowski, J., and Jakubowicz, K., 1972. "Epidermodysplasia Verruciformis as a Model in Studies on the Role of Papovaviruses in Oncogenesis." *Cancer Research*, **32**(3): 583–9. 67

Jacobs, P.A., Tough, I.M., and Wright, D.H., 1963. "Cytogenetic Studies in Burkitt's Lymphoma." *The Lancet*, **282**(7318): 1144–6. 47

Jacomo, V., Kelly, P.J., and Raoult, D., 2002. "Natural History of *Bartonella* Infections (an Exception to Koch's Postulate)." *Clinical and Diagnostic Laboratory Immunology*, **9**(1): 8–18. 18

Javier, R.T. and Butel, J.S., 2008. "The History of Tumor Virology." *Cancer Research*, **68**(19): 7693–706. 15

Johnson, R.J., Floege, J., Couser, W.G., and Alpers, C.E., 1993. "Role of Platelet-derived Growth Factor in Glomerular Disease." *Journal of the American Society of Nephrology*, **4**(2): 119–28. 18

Joint Formulary Committee, 2009. "British National Formulary. 58th ed: Human Papilloma Virus Vaccines." URL `http://bnf.org/bnf/bnf/current/201182.htm`. 75

Joles, J.A., 2008. "Statins and Small GTPases: Koch's Postulates and Chronic Kidney Disease." *Nephrology, Dialysis, Transplantation*, **23**(2): 433–8. 18

Jones, E., MacDonald, I., and Breslow, L., 1958. "A Study of Epidemiologic Factors in Carcinoma of the Uterine Cervix." *American Journal of Obstetrics and Gynecology*, **76**(1): 1–10. 56

Joske, R.A., Leak, P.J., Papadimitriou, J.M., Stanley, N.F., and Walters, M.N., 1966. "Murine Infection with Reovirus. IV. Late Chronic Disease and the Induction of Lymphoma After Reovirus Type 3 Infection." *British Journal of Experimental Pathology*, **47**(4): 337–46. 47

Kant, I., 1997 (1790). *Lectures on Metaphysics*. Cambridge: Cambridge University Press. 148

Karajannis, M.A., Hummel, M., Anagnostopoulos, I., and Stein, H., 1997. "Strict Lymphotropism of Epstein-Barr Virus during Acute Infectious Mononucleosis in Nonimmunocompromised Individuals." *Blood*, **89**(8): 2856–62. 27

Kashgarian, M. and Dunn, J.E., 1970. "The Duration of Intraepithelial and Preclinical Squamous Cell Carcinoma of the Uterine Cervix." *American Journal of Epidemiology*, **92**(4): 211–22. 59

Kassirer, J.P., 1989. "Diagnostic Reasoning." *Annals of Internal Medicine*, **110**(11): 893–900. 19

Kaur, P. and McDougall, J.K., 1988. "Characterization of Primary Human Keratinocytes Transformed by Human Papillomavirus Type 18." *Journal of Virology*, **62**(6): 1917–24. 73

Kennaway, E., 1947. "Cancer of the Penis and Circumcision in Relation to the Incubation Period of Cancer." *British Journal of Cancer*, **1**(4): 335–44. 59

———, 1948. "The Racial and Social Incidence of Cancer of the Uterus." *British Journal of Cancer*, **2**(3): 177–212. 55, 57, 58, 59

Kessler, I.I., 1976. "Human Cervical Cancer as a Venereal Disease." *Cancer Research*, **36**(2 pt 2): 783–91. 60

———, 1977. "Venereal Factors in Human Cervical Cancer. Evidence from Marital Clusters." *Cancer*, **39**(S4): 1912–9. 60

Khanolkar, V.R., 1958. "Cancer in India in Relation to Habits and Customs." In R.W. Raven, editor, *Cancer*, London: Butterworth, volume 3. 272–80. 59

Khoury, M.J., Flanders, W.D., Greenland, S., and Adams, M.J., 1989. "On the Measurement of Susceptibility in Epidemiologic Studies." *American Journal of Epidemiology*, **129**(1): 183–90. 19

Kieff, E. and Rickinson, A.B., 2001. "Epstein-Barr Virus and its Replication." In D.M. Knipe and P.M. Howley, editors, *Field's Virology*, Philadelphia: Lippincott Williams and Company. 4th edition, 2511–73. 27, 28

Kimura, S., Flannery, V.L., Levy, B., and Schaffer, P.A., 1975. "Oncogenic Transformation of Primary Hamster Cells by Herpes Simplex Virus Type 2 (HSV-2) and an HSV-2 Temperature-Sensitive Mutant." *International Journal of Cancer*, **15**(5): 786–98. 62

Kingery, L., 1921. "The Etiology of Common Warts. Their Production in the Second Generation." *Journal of the American Medical Association*, **76**(7): 440–2. 66

Kitcher, P., 1984. "1953 and All That. A Tale of Two Sciences." *The Philosophical Review*, **93**(3): 335–73. 136

———, 1989. "Explanatory Unification and the Causal Structure of the World." In P. Kitcher and W.C. Salmon, editors, *Minnesota Studies in the Philosophy of Science*, Minneapolis, MN: University of Minnesota Press, volume 13. 410–505. 134

Klein, G., 1973. "Summary of Papers Delivered at the Conference on Herpesvirus and Cervical Cancer (Key Biscayne, Florida)." *Cancer Research*, **33**(6): 1557–63. 62, 63, 64

Klug, S.J., Molijn, A., Schopp, B., Holz, B., Iftner, A., Quint, W., Snijders, J.F.P., Petry, K., Kjær, S.K., Munk, C., and Iftner, T., 2008. "Comparison of the Performance of Different HPV Genotyping Methods for Detecting Genital HPV Types." *Journal of Medical Virology*, **80**(7): 1264–74. 75

von Knebel Doeberitz, M., Oltersdorf, T., Schwarz, E., and Gissmann, L., 1988. "Correlation of Modified Human Papilloma Virus Early Gene Expression with Altered Growth Properties in C4-1 Cervical Carcinoma Cells." *Cancer Research*, **48**(13): 3780–6. 73

von Knebel Doeberitz, M., Rittmüller, C., Aengeneyndt, F., Jansen-Dürr, P., and Spitkovsky, D., 1994. "Reversible Repression of Papillomavirus Oncogene Expression in Cervical Carcinoma Cells: Consequences for the Phenotype and E6-p53 and E7-pRb Interactions." *Journal of Virology*, **68**(5): 2811–21. 73

von Knebel Doeberitz, M., Rittmüller, C., zur Hausen, H., and Dürst, M., 1992. "Inhibition of Tumorigenicity of Cervical Cancer Cells in Nude Mice by HPV E6-E7 Anti-Sense RNA." *International Journal of Cancer*, **51**(5): 831–4. 74

Koh, K.K., Ahn, J.Y., Han, S.H., Kim, D.S., Jin, D.K., Kim, H.S., Shin, M., Ahn, T.H., Choi, I.S., and Shin, E.K., 2003. "Pleiotropic Effects of Angiotensin II Receptor Blocker in Hypertensive Patients." *Journal of the American College of Cardiology*, **42**(5): 905–10. 131

Komatsu, T.L., Rivero, E.R.C., de Magalhes, M.H.C.G., and Nunes, F.D., 2005. "Epstein-Barr Virus in Oral Hairy Leukoplakia Scrapes: Identification by PCR." *Brazilian Oral Research*, **19**(4): 317–21. 28

Kondo, K., Ochi, H., Matsumoto, T., Yoshikawa, H., and Kanda, T., 2008. "Modification of Human Papillomavirus-Like Particle Vaccine by Insertion of the Cross-Reactive L2-Epitopes." *Journal of Medical Virology*, **80**(5): 841–6. 75

Koopman, J.S., 1981. "Interaction between Discrete Causes." *American Journal of Epidemiology*, **113**(6): 716–24. 19

Koutsky, L.A., Galloway, D.A., and Holmes, K.K., 1988. "Epidemiology of Genital Human Papillomavirus Infection." *Epidemiologic Reviews*, **10**: 122–63. 69

Krcmár, M., Suchánková, A., Kanka, J., and Vonka, V., 1986. "Prospective Study on the Relationship Between Cervical Neoplasia and Herpes Simplex Type 2 Virus. III. Presence of Herpes Simplex Type-2 Antibody in Sera of Subjects Who Developed Cervical Neoplasia Later in the Study." *International Journal of Cancer*, **38**(2): 161–5. 65, 171, 177

Kruger, D.H., Schneck, P., and Gelderblom, H.R., 2000. "Helmut Ruska and the Visualisation of Viruses." *The Lancet*, **355**(9216): 1713–7. 15

Kuhn, T.S., 1962. "Historical Structure of Scientific Discovery." *Science*, **136**(3518): 760–4. 20, 21, 22, 31

Kumar, S. and Taylor, G., 1973. "The Response to Phytohaemagglutinin (PHA) of Lymphocytes from Cancer Patients." *Journal of Clinical Pathology*, **26**(7): 476–9. 102

Kutinová, L., Vonka, V., and Broucěk, J., 1973. "Increased Oncogenicity and Synthesis of Herpesvirus Antigens in Hamster Cells Exposed to Herpes Simplex Type-2 Virus." *Journal of the National Cancer Institute*, **50**(3): 759–66. 62

Larner, J. and Villar-Palasi, C., 1959. "Enzymes in a Glycogen Storage Myopathy." *Proceedings of the National Academy of Sciences of the United States of America*, **45**(8): 1234–5. 98

Lee, E.S., Locker, J., Nalesnik, M., Reyes, J., Jaffe, R., Alashari, M., Nour, B., Tzakis, A., and Dickman, P.S., 1995. "The Association of Epstein-Barr Virus with Smooth-muscle Tumors Occurring after Organ Transplantation." *The New England Journal of Medicine*, **332**(1): 19–25. 28

Lehtinen, M., Hakama, M., Aaran, R.K., Aromaa, A., Knekt, P., Leinikki, P., Maatela, J., Peto, R., and Teppo, L., 1992. "Herpes Simplex Virus Type 2 Infection and Cervical Cancer: a Prospective Study of 12 Years of Follow-up in Finland." *Cancer Causes & Control*, **3**(4): 333–8. 65

Lejeune, J., Gautier, M., and Turpin, R., 1959. "Étude des Chromosomes Somatiques de Neuf Enfants Mongoliens [Study of the Somatic Chromosomes of Nine Mongoloid Children]." *Comptes Rendus Hebdomadaires Des Séances De l'Académie Des Sciences*, **248**(11): 1721–2. 155

Leplus, R., Debray, J., Pinet, J., and Bernhard, W., 1961. "Lésions Nucléaires Décelées au Microscope Électronique dans des Cellules de Lymphomes Malins chez l'Homme. [Nuclear Lesions Revealed by the Electron Microscope in the Cells of Malignant Lymphomas in Man]." *Comptes Rendus Hebdomadaires Des Séances De l'Académie Des Sciences*, **253**(12): 2788–90. 43

Levy, J.A., Tanabe, E., and Curnen, E.C., 1968. "Occurrence of Reovirus Antibodies in Health African Children and in Children with Burkitt's Lymphoma." *Cancer*, **21**(1): 53–7. 47

Lewandowsky, F. and Lutz, W., 1922. "Ein Fall einer bisher nicht Beschriebenen Hauterkrankung (Epidermodysplasia Verruciformis)[A Case of a Hitherto Undescribed Skin Disease (Epidermodysplasia Verruciformis)]." *Archiv für Dermatologie und Syphilis*, **141**(2): 193–203. 67

Lewis, D., 1979. "Counterfactual Dependence and Time's Arrow." In *Philosophical Papers: Volume II*, Oxford: Oxford University Press. 32–66. 108

———, 1986. *Philosophical Papers: Volume II*. Oxford: Oxford University Press. 108, 146

———, 2000. *Counterfactuals*. Blackwell. 82, 108

Liao, J.K. and Laufs, U., 2005. "Pleiotropic Effects of Statins." *Annual Review of Pharmacology and Toxicology*, **45**: 89–118. 131

Lilienfeld, A.M., 1980. *Times, Places, and Persons: Aspects of the History of Epidemiology*. Baltimore, Maryland: John's Hopkins University Press. 17

Lilienfeld, D.E. and Lilienfeld, A.M., 1977. "Epidemiology: A Retrospective Study." *American Journal of Epidemiology*, **106**(6): 445–59. 17

Lipton, R. and Odegaard, T., 2005. "Causal Thinking and Causal Language in Epidemiology: It's in the Details." *Epidemiologic Perspectives & Innovations*, **2**: 8. 77

Lombard, H. and Potter, E., 1950. "Environmental Factors in the Etiology of Cancer." *International Journal of Cancer*, **6**(6): 1325–33. 57

Lothe, F. and Somers, K., 1960. "Secondary Tumors of the Heart and Pericardium in Uganda Africans." *Archives of Pathology*, **69**(2): 158–67. 36

Lowy, D. and Howley, P., 2001. "Papillomaviruses." In *Field's Virology*, Philadelphia: Lippincott Williams and Company. 4th edition, 2231–64. 54

Lucké, B., 1934. "A Neoplastic Disease of the Kidney of the Frog, (*Rana pipiens*)." *American Journal of Cancer*, **20**: 352–79. 16, 127

———, 1938a. "Carcinoma in the Leopard Frog: Its Probable Causation by a Virus." *The Journal of Experimental Medicine*, **68**(4): 457–68. 16, 127

———, 1938b. "Carcinoma of the Kidney in the Leopard Frog: The Occurrence and Significance of Metastasis." *American Journal of Cancer*, **34**(1): 15–30. 16, 127

———, 1939. "Characteristics of Frog Carcinoma in Tissue Culture." *The Journal of Experimental Medicine*, **70**(3): 269–76. 16, 127

———, 1952. "Kidney Carcinoma in the Leopard Frog: A Virus Tumor." *Annals of the New York Academy of Sciences*, **54**(Viruses as Causative Agents in Cancer): 1093–109. 16, 127

Lyell, A. and Miles, J.A.R., 1951. "The Myrmecia: A Study of Inclusion Bodies in Warts." *British Medical Journal*, **1**(4712): 912–5. 66

Machamer, P., 2004. "Activities and Causation: The Metaphysics and Epistemology of Mechanisms." *International Studies in the Philosophy of Science*, **18**(1): 27–39. 110, 111, 127, 133

Machamer, P., Darden, L., and Craver, C.F., 2000. "Thinking about Mechanisms." *Philosophy of Science*, **67**(1): 1–25. 110, 112, 113, 114, 115, 137, 182

Mackie, J.L., 1974. *The Cement of the Universe: A Study of Causation*. Oxford: Oxford University Press. 19, 108

MacMahon, B. and Trichopoulos, D., 1996. *Epidemiology, Principles and Methods*. Boston: Little Brown. 17, 19

MacMahon, E., Charache, P., Glass, D., Mann, R.B., Becker, P.S., Hayward, D., McArthur, J., and Ambinder, R.F., 1991. "Epstein-Barr Virus in AIDS-related Primary Central Nervous System Lymphoma." *The Lancet*, **338**(8773): 969–73. 28

MacNab, J.C., 1974. "Transformation of Rat Embryo Cells by Temperature-Sensitive Mutants of Herpes Simplex Virus." *The Journal of General Virology*, **24**(1): 143–53. 62

Macville, M., Schröck, E., Padilla-Nash, H., Keck, C., Ghadimi, B.M., Zimonjic, D., Popescu, N., and Ried, T., 1999. "Comprehensive and Definitive Molecular Cytogenetic Characterization of HeLa Cells by Spectral Karyotyping." *Cancer Research*, **59**(1): 141–50. 71

Magrath, I., 2002. "The Non-Hodgkin's Lymphomas in Children." In R. Souhami, I. Tannock, P. Hohenberger, and J. Horiot, editors, *The Oxford Textbook of Oncology*, Oxford: Oxford University Press. 2nd edition, 2378–418. 26, 27

MARA, 2010. "MARA (Mapping Malaria Risk in Africa) Website." URL `http://www.mara.org.za/mapsinfo.htm`. 41

Marek, J., 1907. "Multiple Nervenentzündung (Polyneuritis) bei Hühnern [Multiple Neuritis (Polyneuritis) in Chickens]." *Deutsche Tierärztliche Wochenschrift*, **15**: 417–21. 15

Martin, C.E., 1967. "Epidemiology of Cancer of the Cervix. II. Marital and Coital Factors in Cervical Cancer." *American Journal of Public Health and the Nation's Health*, **57**(5): 803–14. 56, 58

Martini, F., Corallini, A., Balatti, V., Sabbioni, S., Pancaldi, C., and Tognon, M., 2007. "Simian Virus 40 in Humans." *Infectious Agents and Cancer*, **2**: 13. 17

Massa, S., Simeone, P., Muller, A., Benvenuto, E., Venuti, A., and Franconi, R., 2008. "Antitumor Activity of DNA Vaccines based on the Human Papillomavirus-16 E7 Protein Genetically Fused to a Plant Virus Coat Protein." *Human Gene Therapy*, **19**(4): 354–64. 75

Maxwell, G.M., 1954. "Twelve Cases of Lymphoblastomata in Children." *Archives of Disease in Childhood*, **29**(144): 155–9. 38

Mayrand, M., Duarte-Franco, E., Coutlée, F., Rodrigues, I., Walter, S.D., Ratnam, S., and Franco, E.L., 2006. "Randomized Controlled Trial of Human Papillomavirus Testing Versus Pap Cytology in the Primary Screening for Cervical Cancer Precursors: Design, Methods and Preliminary Accrual Results of the Canadian Cervical Cancer Screening Trial (CCCAST)." *International Journal of Cancer*, **119**(3): 615–23. 75

Mayrand, M., Duarte-Franco, E., Rodrigues, I., Walter, S.D., Hanley, J., Ferenczy, A., Ratnam, S., Coutlée, F., and Franco, E.L., 2007. "Human Papillomavirus DNA versus Papanicolaou Screening Tests for Cervical Cancer." *The New England Journal of Medicine*, **357**(16): 1579–88. 75

Mazzarello, P., 2004. "Il Potere Esplicativo delle Regole: Il caso dei Postulati di Koch [The Explicative Power of the Rules: the case of Koch's postulates]." *Medicina Nei Secoli*, **16**(2): 293–316. 18

McAleer, W.J., Buynak, E.B., Maigetter, R.Z., Wampler, D.E., Miller, W.J., and Hilleman, M.R., 1984. "Human Hepatitis B Vaccine from Recombinant Yeast." *Nature*, **307**(5947): 178–80. 104

McArdle, B., 1951. "Myopathy Due to a Defect in Muscle Glycogen Breakdown." *Clinical Science*, **10**(1): 13–33. 97, 99, 157

McCance, D.J., 1998. "Human Papillomaviruses and Cervical Cancer." *Journal of Medical Microbiology*, **47**(5): 371–3. 54

———, 2004. "Papillomaviruses." In A.J. Zuckerman, J.E. Banatvala, J.R. Pattison, P. Griffiths, and B. Schoub, editors, *Principles and Practice of Clinical Virology*, Chichester: John Wiley & Sons. 5th edition, 661–74. 54, 73

McClain, K.L., Leach, C.T., Jenson, H.B., Joshi, V.V., Pollock, B.H., Parmley, R.T., DiCarlo, F.J., Chadwick, E.G., and Murphy, S.B., 1995. "Association of Epstein-Barr Virus with Leiomyosarcomas in Children with AIDS." *The New England Journal of Medicine*, **332**(1): 12–8. 28

McGraw, D.J., 1989. "Essays of Robert Koch." *Isis*, **80**(3): 542–3. 18

McKay Illari, P., 2010. "Mechanistic Evidence: Disambiguating the Russo-Williamson Thesis." Unpublished MS draft. 85, 152, 164

Meisels, A., Bégin, R., and Schneider, V., 1977. "Dysplasias of Uterine Cervix: Epidemiological Aspects: Role of Age at First Coitus and Use of Oral Contraceptives." *Cancer*, **40**(6): 3076–81. 58

Meléndez, L.V., Daniel, M.D., García, F.G., Fraser, C.E., Hunt, R.D., and King, N.W., 1969a. "*Herpesvirus saimiri*. I. Further Characterization Studies of a New Virus from the Squirrel Monkey." *Laboratory Animal Care*, **19**(3): 372–7. 17

Meléndez, L.V., Daniel, M.D., Hunt, R.D., and García, F.G., 1968. "An Apparently new Herpesvirus from Primary Kidney Cultures of the Squirrel Monkey (*Saimiri sciureus*)." *Laboratory Animal Care*, **18**(3): 374–81. 17

Meléndez, L.V., Hunt, R.D., Daniel, M.D., García, F.G., and Fraser, C.E., 1969b. "*Herpesvirus saimiri*. II. Experimentally Induced Malignant Lymphoma in Primates." *Laboratory Animal Care*, **19**(3): 378–86. 17

Meléndez, L.V., Hunt, R.D., King, N.W., Barahona, H.H., Daniel, M.D., Fraser, C.E., and García, F.G., 1972. "*Herpesvirus ateles*, a New Lymphoma Virus of Monkeys." *Nature: New Biology*, **235**(58): 182–4. 17

Melnick, J.L. and Adam, E., 1978. "Epidemiological Approaches to Determining Whether Herpesvirus Is the Etiological Agent of Cervical Cancer." *Progress in Experimental Tumor Research*, **21**(1): 49–69. 64

Melnick, J.L., Adam, E., and Rawls, W.E., 1974. "The Causative Role of Herpesvirus Type 2 in Cervical Cancer." *Cancer*, **34**(Suppl 4): 1375–85. 63

Melnick, J.L., Bunting, H., Banfield, W.G., Strauss, M.J., and Gaylord, W.H., 1952. "Electron Microscopy of Viruses of Human Papilloma, Molluscum Contagiosum, and Vaccinia, Including Observations on the Formation of Virus Within the Cell." *Annals of the New York Academy of Sciences*, **54**(Viruses as Causative Agents in Cancer): 1214–25. 66

Mendelson, C.G. and Kligman, A.M., 1961. "Isolation of Wart Virus in Tissue Culture. Successful Reinoculation into Humans." *Archives of Dermatology*, **83**(4): 559–62. 67

Mengert, W.F., Brown, W.W., and Cobb, S.W., 1949. "Modern Concepts of Cervical Carcinoma." *Obstetrical and Gynecological Survey*, **4**(3): 315–7. 55

Menzies, P. and Price, H., 1993. "Causation as a Secondary Quality." *The British Journal for the Philosophy of Science*, **44**(2): 187–203. 109

Miettinen, O.S., 1982. "Causal and Preventive Interdependence. Elementary Principles." *Scandinavian Journal of Work, Environment & Health*, **8**(3): 159–68. 19

Miles, C.P. and O'Neill, F., 1967. "Chromosome Studies of 8 *in vitro* Lines of Burkitt's Lymphoma." *Cancer Research*, **27**(2): 392–402. 49

Miller, G., 1974. "The Oncogenicity of Epstein-Barr Virus." *The Journal of Infectious Diseases*, **130**(2): 187–205. 47

Minson, A.C., Thouless, M.E., Eglin, R.P., and Darby, G., 1976. "The Detection of Virus DNA Sequences in a Herpes Type 2 Transformed Hamster Cell Line (333-8-9)." *International Journal of Cancer*, **17**(4): 493–500. 62

de Moliaure, J.B.P., 1673. *Le Malade Imaginaire*. 129

Moloney, J.B., 1962. "The Murine Leukemias." *Federation Proceedings*, **21**(1): 19–31. 17

Mommaerts, W.F., Illingworth, B., Pearson, C.M., Guillory, R.J., and Seraydarian, K., 1959. "A Functional Disorder of Muscle Associated with the Absence of Phosphorylase." *Proceedings of the National Academy of Sciences of the United States of America*, **45**(6): 791–7. 97, 98

Moody, C.A. and Laimins, L.A., 2010. "Human Papillomavirus Oncoproteins: Pathways to Transformation." *Nature Reviews Cancer*, **10**(8): 550–60. 73

Moore, D.H., Stone, R.S., Shope, R.E., and Gelber, D., 1959. "Ultrastructure and Site of Formation of Rabbit Papilloma Virus." *Proceedings of the Society for Experimental Biology and Medicine*, **101**(3): 575–8. 68

Morgan, H.R. and Balduzzi, P.C., 1964. "Propagation of an Intranuclear Inclusion-Forming Agent from Human Condyloma Acuminatum." *Proceedings of the National Academy of Sciences of the United States of America*, **52**(6): 1561–4. 67

Morris, J. and Meigs, J., 1950. "Carcinoma of the Cervix; Statistical Evaluation of 1,938 Cases and Results of Treatment." *Surgery, Gynecology & Obstetrics*, **90**(2): 135–50. 58

Münger, K., Phelps, W.C., Bubb, V., Howley, P.M., and Schlegel, R., 1989. "The E6 and E7 Genes of the Human Papillomavirus Type 16 Together Are Necessary and Sufficient for Transformation of Primary Human Keratinocytes." *Journal of Virology*, **63**(10): 4417–21. 73

Muñoz, N., 1973. "Effect of Herpesvirus type 2 and Hormonal Imbalance on the Uterine Cervix of the Mouse." *Cancer Research*, **33**(6): 1504–8. 62

Muñoz, N., Bosch, F.X., de Sanjosé, S., Tafur, L., Izarzugaza, I., Gili, M., Viladiu, P., Navarro, C., Martos, C., and Ascunce, N., 1992. "The Causal Link Between Human Papillomavirus and Invasive Cervical Cancer: A Population-Based Case-Control Study in Colombia and Spain." *International Journal of Cancer*, **52**(5): 743–9. 74

Munube, G.M. and Bell, T.M., 1967. "Isolation of Echovirus Type 11 from Two Cases of Burkitt's Tumour and Three Cases of Other Tumours." *International Journal of Cancer*, **2**(6): 613–8. 48

Murray, J.F. and Brandt, F.A., 1951. "Histoplasmosis and Malignant Lymphoma." *The American Journal of Pathology*, **27**(5): 783–99. 35, 37

Naegele, R.F. and Granoff, A., 1977. "Viruses and Renal Carcinoma of *Rana pipiens*. XV. The Presence of Virus-associated Membrane Antigen(s) on Lucké Tumor Cells." *International Journal of Cancer*, **19**(3): 414–8. 17

Nahmias, A.J., Naib, Z.M., Josey, W.E., Murphy, F.A., and Luce, C.F., 1970. "Sarcomas After Inoculation of Newborn Hamsters with Herpes Virus Hominis Type 2 Strains." *Proceedings of the Society for Experimental Biology and Medicine*, **134**(4): 1065–9. 62

Naib, Z.M., Nahmias, A.J., and Josey, W.E., 1966. "Cytology and Histopathology of Cervical Herpes Simplex Infection." *Cancer*, **19**(7): 1026–31. 60

Naib, Z.M., Nahmias, A.J., Josey, W.E., and Zaki, S.A., 1973. "Relation of Cytohistopathology of Genital Herpesvirus Infection to Cervical Anaplasia." *Cancer Research*, **33**(6): 1452–63. 64

Niedobitek, G. and Young, L.S., 1994. "Epstein-Barr Virus Persistence and Virus-associated Tumours." *The Lancet*, **343**(8893): 333–5. 27

Nigam, R., 1954. "Tumours of the Jaw." *Indian Journal of Surgery*, **16**(4): 279–307. 38

Noyes, W.F., 1965. "Studies on the Human Wart Virus. II. Changes in Primary Human Cell Cultures." *Virology*, **25**(3): 358–63. 67

O'Conor, G.T., 1961. "Malignant Lymphoma in African Children. II. A Pathological Entity." *Cancer*, **14**(2): 270–83. 33, 34, 37, 38

O'Conor, G.T. and Davies, J.N.P., 1960. "Malignant Tumors in African Children. With Special Reference to Malignant Lymphoma." *The Journal of Pediatrics*, **56**(4): 526–35. 32, 33, 35, 36

Okagaki, T., Twiggs, L.B., Zachow, K.R., Clark, B.A., Ostrow, R.S., and Faras, A.J., 1983. "Identification of Human Papillomavirus DNA in Cervical and Vaginal Intraepithelial Neoplasia with Molecularly Cloned Virus-Specific DNA Probes." *International Journal of Gynecological Pathology: Official Journal of the International Society of Gynecological Pathologists*, **2**(2): 153–9. 71

Okochi, K. and Murakami, S., 1968. "Observations on Australia Antigen in Japanese." *Vox Sanguinis*, **15**(5): 374–85. 101, 102

Olson, C., Pamukcu, A.M., Brobst, D.F., Kowalczyk, T., Satter, E.J., and Price, J.M., 1959. "A Urinary Bladder Tumor Induced by a Bovine Cutaneous Papilloma Agent." *Cancer Research*, **19**(7): 779–82. 68

Oppenheimer, R., 1956. "Analogy in Science." *American Psychologist*, **11**(3): 127–35. 166

Oriel, J., 1971. "Natural History of Genital Warts." *British Journal of Venereal Disease*, **47**(1): 1–13. 66, 67

Oriel, J. and Almeida, J., 1970. "Demonstration of Virus Particles in Human Genital Warts." *The British Journal of Venereal Diseases*, **46**(1): 37–42. 67, 68, 69

Oroszlan, S. and Rich, M.A., 1964. "Human Wart Virus: *in vitro* Cultivation." *Science*, **146**(3643): 531–3. 67

Orsós, S., 1958. "Primary Lymphosarcoma of the Gingivae." *Oral Surgery, Oral Medicine and Oral Pathology*, **11**(4): 426–30. 38

Orth, G., Favre, M., and Croissant, O., 1977. "Characterization of a New Type of Human Papillomavirus That Causes Skin Warts." *Journal of Virology*, **24**(1): 108–20. 70, 72

Orth, G., Favre, M., Jablonska, S., Brylak, K., and Croissant, O., 1978a. "Viral Sequences Related to a Human Skin Papillomavirus in Genital Warts." *Nature*, **275**(5678): 334–6. 70

Orth, G., Jablonska, S., Favre, M., Croissant, O., Jarzabek-Chorzelska, M., and Rzesa, G., 1978b. "Characterization of Two Types of Human Papillomaviruses in Lesions of Epidermodysplasia Verruciformis." *Proceedings of the National Academy of Sciences of the United States of America*, **75**(3): 1537–41. 67, 70, 72

Orth, G., Jablonska, S., Jarzabek-Chorzelska, M., Obalek, S., Rzesa, G., Favre, M., and Croissant, O., 1979. "Characteristics of the Lesions and Risk of Malignant Conversion Associated with the Type of Human Papillomavirus Involved in Epidermodysplasia Verruciformis." *Cancer Research*, **39**(3): 1074–82. 70

Orth, G., Jablonska, S., Favre, M., Croissant, O., Obalek, S., Jarzabek-Chorzelska, M., and Jibard, N., 1981. "Identification of Papillomaviruses in Butchers' Warts." *Journal of Investigative Dermatology*, **76**(2): 97–102. 71

O'Shaughnessy, J., Osborne, C., Pippen, J.E., Yoffe, M., Patt, D., Rocha, C., Koo, I.C., Sherman, B.M., and Bradley, C., 2011. "Iniparib plus Chemotherapy in Metastatic Triple-Negative Breast Cancer." *New England Journal of Medicine*, **364**(3): 205–14. 85

Ostrow, R.S., Krzyzek, R., Pass, F., and Faras, A.J., 1981. "Identification of a Novel Human Papilloma Virus in Cutaneous Warts of Meathandlers." *Virology*, **108**(1): 21–7. 71

Paavonen, J., 2007. "Human Papillomavirus Infection and the Development of Cervical Cancer and Related Genital Neoplasias." *International Journal of Infectious Diseases*, **11**(Suppl 2): S3–9. 54

Pagano, J.S., 1975. "Diseases and Mechanisms of Persistent DNA Virus Infection: Latency and Cellular Transformation." *The Journal of Infectious Diseases*, **132**(2): 209–23. 65

Pallesen, G., Hamilton-Dutoit, S.J., and Zhou, X., 1993. "The Association of Epstein-Barr Virus (EBV) with T cell Lymphoproliferations and Hodgkin's Disease: Two New Developments in the EBV Field." *Advances in Cancer Research*, **62**: 179–239. 28

Palmer, A.E., London, W.T., Nahmias, A.J., Naib, Z.M., Tunca, J., Fuccillo, D.A., Ellenberg, J.H., and Sever, J.L., 1976. "A Preliminary Report on Investigation of Oncogenic Potential of Herpes Simplex Virus Type 2 in Cebus Monkeys." *Cancer Research*, **36**(2 pt 2): 807–9. 62

Papadimitriou, J.M., 1966. "Electron Microscopic Findings of a Murine Lymphoma Associated with Reovirus Type 3 Infection." *Proceedings of the Society for Experimental Biology and Medicine*, **121**(1): 93–6. 17

Pappenheimer, A.M., Dunn, L.C., and Cone, V., 1929a. "Studies on Fowl Paralysis (Neurolymphomatosis Gallinarum): I. Clinical Features and Pathology." *Journal of Experimental Medicine*, **49**(1): 63–86. 15

Pappenheimer, A.M., Dunn, L.C., and Seidlin, S.M., 1929b. "Studies on Fowl Paralysis (Neurolymphomatosis Gallinarum): II. Transmission Experiments." *Journal of Experimental Medicine*, **49**(1): 87–102. 15

Parascandola, M. and Weed, D.L., 2001. "Causation in Epidemiology." *Journal of Epidemiology and Community Health*, **55**(12): 905–12. 79

Parker, J.D. and Banatvala, J.E., 1967. "Herpes Genitalis; Clinical and Virological Studies." *The British Journal of Venereal Diseases*, **43**(3): 212–6. 60

Parker, L., Baker, E., and Stanley, N., 1965. "The Isolation of Reovirus Type 3 from Mosquitoes and a Sentinel Infant Mouse." *Australian Journal of Experimental Biology and Medical Science*, **43**(2): 167–70. 47

Parkin, D.M., Pisani, P., and Ferlay, J., 1993. "Estimates of the Worldwide Incidence of Eighteen Major Cancers in 1985." *International Journal of Cancer*, **54**(4): 594–606. 54

Parkin, D.M., Stjernswrd, J., and Muir, C.S., 1984. "Estimates of the Worldwide Frequency of Twelve Major Cancers." *Bulletin of the World Health Organization*, **62**(2): 163–82. 54

Pass, F., Reissig, M., Shah, K.V., Eisinger, M., and Orth, G., 1977. "Identification of an Immunologically Distinct Papillomavirus from Lesions of Epidermodysplasia Verruciformis." *Journal of the National Cancer Institute*, **59**(4): 1107–12. 70

Passey, R.D., Dmochowski, L., Astbury, W.T., Reed, R., and Johnson, P., 1950. "Electron Microscope Studies of Normal and Malignant Tissues of High- and Low- Breast-Cancer Strains of Mice." *Nature*, **165**(4186): 107. 16

Payet, M., Camain, R., and Pene, P., 1956. "Le Cancer Primitif du Foie, Etude Critique a propos de 240 Cas. [Primary Cancer of the Liver: Critical Study of 240 Cases]." *Revue Internationale D'hépatologie*, **6**(1): 1–86. 102

Pearl, J., 1993. "Graphical models, Causality, and Intervention." *Statistical Science*, **8**(3): 266–73. 19

———, 1987. "Evidential Reasoning Using Stochastic Simulation of Causal Models." *Artificial Intelligence*, **32**(2): 245–57. 178

———, 1995. "Causal Diagrams for Empirical Research." *Biometrika*, **82**(4): 669–88. 19

———, 2000. *Casuality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press. 19

Pearson, C., Rimer, D., and Mommaerts, W., 1961. "A Metabolic Myopathy Due to Absence of Muscle Phosphorylase." *The American Journal of Medicine*, **30**(4): 502–17. 97

Pecoraro, G., Lee, M., Morgan, D., and Defendi, V., 1991. "Evolution of *in vitro* Transformation and Tumorigenesis of HPV16 and HPV18 Immortalized Primary Cervical Epithelial Cells." *The American Journal of Pathology*, **138**(1): 1–8. 73

Peers, F.G. and Linsell, C.A., 1973. "Dietary Aflatoxins and Liver Cancer–A Population Based Study in Kenya." *British Journal of Cancer*, **27**(6): 473–84. 103

Pender, M.P., 2009. "Preventing and Curing Multiple Sclerosis by Controlling Epstein-Barr Virus Infection." *Autoimmunity Reviews*, **8**(7): 563–8. 28

Pereira, M.S., Pereira, H.G., and Clarke, S.K.R., 1965. "Human Adenovirus Type 31 a New Serotype with Oncogenic Properties." *The Lancet*, **285**(7375): 21–3. 17

Pfister, H., Nrnberger, F., Gissmann, L., and zur Hausen, H., 1981. "Characterization of a Human Papillomavirus from Epidermodysplasia Verruciformis Lesions of a Patient from Upper-Volta." *International Journal of Cancer*, **27**(5): 645–50. 71

Phan-Ngoc-Duong and Dao-Duc-Hoanh, 1961. "Contributions à L'étude De La Méthode Combinée Chirugie-Curiethérapie Dans Le Traitement Des Cancers Du Maxillaire Supérieur [Contributions to the Study of Combined Surgery and Brachytherapy in the Treatment of Cancers of the Maxilla]." *Annals of the Institute of Cancer of Viet-nam*, **Unknown Volume**: 47–52. 38

Picard, B., Garcia, J.S., Gouriou, S., Duriez, P., Brahimi, N., Bingen, E., Elion, J., and Denamur, E., 1999. "The Link between Phylogeny and Virulence in *Escherichia coli* Extraintestinal Infection." *Infection and Immunity*, **67**(2): 546–53. 170

Pirisi, L., Yasumoto, S., Feller, M., Doniger, J., and DiPaolo, J.A., 1987. "Transformation of Human Fibroblasts and Keratinocytes with Human Papillomavirus Type 16 DNA." *Journal of Virology*, **61**(4): 1061–6. 73

Pitot, H.C., 1983. "Landmark Perspective: The Rous Sarcoma Virus." *Journal of the American Medical Association*, **250**(11): 1447–9. 16

Plaut, A. and Kohn-Speyer, A., 1947. "The Carcinogenic Action of Smegma." *Science*, **105**(2728): 391–2. 59

Plummer, G., 1973. "A Review of the Identification and Titration of Antibodies to Herpes Simplex Viruses Type 1 and Type 2 in Human Sera." *Cancer Research*, **33**(6): 1469–76. 64

Powell, N.G., Hibbitts, S.J., Boyde, A.M., Newcombe, R.G., Tristram, A.J., and Fiander, A.N., 2010. "The Risk of Cervical Cancer Associated with Specific Types of Human Papillomavirus: A Case-control Study in a UK Population." *International Journal of Cancer*, **Pre-print**. 74

Pratt-Thomas, H.R., Heins, H.C., Latham, E., Dennis, E.J., and McIver, F.A., 1955. "The Carcinogenic Effect of Human Smegma: An Experimental Study. I. Preliminary Report." *Cancer*, **9**(4): 671–80. 59

Price, H., 1991. "Agency and Probabilistic Causality." *The British Journal for the Philosophy of Science*, **42**(2): 157–76. 82

Psillos, S., 2002. *Causation and Explanation*. Chesham: Acumen Publishing. 108, 110

Pulvertaft, J.V., 1964. "Cytology of Burkitt's Tumour (African Lymphoma)." *The Lancet*, **283**(7327): 238–40. 35, 45, 46

Quinlivan, R., Beynon, R.J., and Martinuzzi, A., 2009. "Pharmacological and Nutritional Treatment for McArdle Disease (Glycogen Storage Disease Type V)." *Cochrane Database of Systematic Reviews (Online)*, **3**: CD003458. 97

Rapp, F. and Falk, L., 1964. "Study of Virulence and Tumorigenicity of Variants of Herpes Simplex Virus." *Proceedings of the Society for Experimental Biology and Medicine*, **116**(6): 361–5. 62

Rauscher, F.J., 1962. "A Virus-induced Disease of Mice Characterized by Erythrocytopoiesis and Lymphoid Leukemia." *Journal of the National Cancer Institute*, **29**(3): 515–43. 17

Rawls, W.E., Adam, E., and Melnick, J.L., 1973. "An Analysis of Seroepidemiological Studies of Herpesvirus Type 2 and Carcinoma of the Cervix." *Cancer Research*, **33**(6): 1477–82. 61, 64

Rawls, W.E., Laurel, D., Melnick, J.L., Glicksman, J.M., and Kaufman, R.H., 1968a. "A Search for Viruses in Smegma, Premalignant and Early Malignant Cervical Tissues. The Isolation of Herpesviruses with Distinct Antigenic Properties." *American Journal of Epidemiology*, **87**(3): 647–55. 60

Rawls, W.E., Tompkins, W.A.F., Figueroa, M.E., and Melnick, J.L., 1968b. "Herpesvirus Type 2: Association with Carcinoma of the Cervix." *Science*, **161**(847): 1255–6. 61

Rawls, W.E., Tompkins, W.A.F., and Melnick, J.L., 1969. "Serological Association between Herpesvirus Type 2 and Cervical Cancer." *Bacteriological Proceedings*, **1969**: 178. 61

Reagan, R.L., Delaha, E.C., Stewart, M.T., and Brueckner, A.L., 1954. "Electron Microscopic Studies of the Bivins Papilloma Virus." *American Journal of Veterinary Research*, **15**(55): 323–4. 66

Reichenbach, H., 1949. *The Theory of Probability: An Inquiry into the Logical and Mathematical Foundations of the Calculus of Probability*. Berkeley, CA: University of California Press. 175

———, 1956. *The Direction of Time*. Berkeley, CA: University of Los Angeles Press. 133

Rickinson, A.B. and Kieff, E., 2001. "Epstein-Barr Virus." In D.M. Knipe and P.M. Howley, editors, *Field's Virology*, Philadelphia: Lippincott Williams and Company. 2575–627. 27

Rivers, T.M., 1937. "Viruses and Koch's Postulates." *Journal of Bacteriology*, **33**(1): 1–12. 18

Robins, J.M., 2001. "Data, Design, and Background Knowledge in Etiologic Inference." *Epidemiology*, **12**(3): 313–20. 19

Robins, J.M. and Greenland, S., 1992. "Identifiability and Exchangeability for Direct and Indirect Effects." *Epidemiology*, **3**(2): 143–155. 19

Røjel, J., 1953. "The Interrelation Between Uterine Cancer and Syphilis; a Pathodemographic Study." *Acta Pathologica Et Microbiologica Scandinavica Supplementum*, **97**: 1–82. 58

Romney, S.L., Palan, P.R., Duttagupta, C., Wassertheil-Smoller, S., Wylie, J., Miller, G., Slagle, N.S., and Lucido, D., 1981. "Retinoids and the Prevention of Cervical Dysplasias." *American Journal of Obstetrics and Gynecology*, **141**(8): 890–4. 59

Ronco, G., Giorgi-Rossi, P., Carozzi, F., Confortini, M., Palma, P.D., Mistro, A.D., Gillio-Tos, A., Minucci, D., Naldoni, C., Rizzolo, R., Schincaglia, P., Volante, R., Zappa, M., Zorzi, M., Cuzick, J., Segnan, N., and behalf of the New Technologies for Cervical Cancer Screening Working Group, O., 2008. "Results at Recruitment From a Randomized Controlled Trial Comparing Human Papillomavirus Testing Alone With Conventional Cytology as the Primary Cervical Cancer Screening Test." *Journal of the National Cancer Institute*, **100**(7): 492–501. 75

Rothman, K.J., 1974. "Synergy and Antagonism in Cause-effect Relationships." *American Journal of Epidemiology*, **99**(6): 385–8. 19

———, 1976a. "Causes." *American Journal of Epidemiology*, **104**(6): 587–92. 19

———, 1976b. "The Estimation of Synergy or Antagonism." *American Journal of Epidemiology*, **103**(5): 506–11. 19

Rothman, K.J., Greenland, S., and Lash, T.L., 2008. *Modern Epidemiology*. Philadelphia, PA: Lippincott Williams & Wilkins, 3rd edition. 81

Rotkin, I.D., 1967. "Epidemiology of Cancer of the Cervix. III. Sexual Characteristics of a Cervical Cancer Population." *American Journal of Public Health and the Nation's Health*, **57**(5): 815–29. 60

———, 1973. "A Comparison Review of Key Epidemiological Studies in Cervical Cancer Related to Current Searches for Transmissible Agents." *Cancer Research*, **33**(6): 1353–67. 61, 64

Rous, P., 1911a. "A Sarcoma of the Fowl Transmissible by an Agent Separable from the Tumor Cells." *The Journal of Experimental Medicine*, **13**(4): 397–411. 16

———, 1911b. "Transmission of a Malignant New Growth by Means of a Cell-Free Filtrate." *Journal of the American Medical Association*, **56**(3): 198–9. 16

Rous, P. and Beard, J.W., 1934. "Carcinomatous Changes in Virus-Induced Papillomas of the Skin of the Rabbit." *Proceedings of the Society for Experimental Biology and Medicine*, **32**: 578–80. 16, 68

———, 1935. "The Progression to Carcinoma of Virus Induced Rabbit Papilloma (Shope)." *Journal of Experimental Medicine*, **62**: 523–48. 68

Rous, P. and Friedewald, W.F., 1944. "The Effect of Chemical Carcinogens on Virus-induced Rabbit Papillomas." *Journal of Experimental Medicine*, **79**: 511–38. 68

Rous, P. and Kidd, J.G., 1938. "The Carcinogenic Effect of a Papilloma Virus on the Tarred Skin of Rabbits: I. Description of the Phenomenon." *Journal of Experimental Biology*, **67**: 399–428. 68

Rowson, K.E. and Mahy, B.W., 1967. "Human Papova (Wart) Virus." *Bacteriological Reviews*, **31**(2): 110–31. 66

Rubin, D.B., 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology*, **66**(5): 688–701. 19

———, 1990. "Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies." *Statistical Science*, **5**(4): 472–80. 19

Ruiter, M. and van Mullem, P.J., 1970. "Behavior of Virus in Malignant Degeneration of Skin Lesion in Epidermodysplasia Verruciformis." *The Journal of Investigative Dermatology*, **54**(4): 324–31. 67

Ruska, E., 1986. "Ernst Ruska - Autobiography (Nobel Lectures, Physics 1986)." In T. Frängsmyr and G. Ekspång, editors, *Nobel Lectures, Physics 1981-1990*, Singapore: World Scientific. 15

Russo, F. and Williamson, J., 2007. "Interpreting Causality in the Health Sciences." *International Studies in the Philosophy of Science*, **21**(2): 157–70. 26, 77, 78, 79, 80, 81, 82, 83, 84, 86, 149

———, 2010. "Generic versus Single-case Causality: the Case of Autopsy." Unpublished MS draft. 80, 90, 99, 181

Sackett, D.L., Rosenberg, W.M.C., Gray, J.A.M., Haynes, R.B., and Richardson, W.S., 1996. "Evidence Based Medicine: What It Is and What It Isn't." *British Medical Journal*, **312**(7023): 71–2. 84

Salman, I. and Darlington, C.G., 1944. "Rare (Unusual) Tumors of the Jaws." *American Journal of Orthodontics*, **30**(12): 725–49. 38

Salmon, W.C., 1977. "Objectively Homogeneous Reference Classes." *Synthese*, **36**(4): 399–414. 175

———, 1980. "Causality: Production and Propagation." *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, **1980**: 49–69. 133

———, 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton, NJ: Princeton University Press. 110, 133, 134, 135

———, 1989. "Four Decades of Scientific Explanation." In P. Kitcher and W.C. Salmon, editors, *Minnesota Studies in the Philosophy of Science*, Minneapolis, MN: Minnesota University Press, volume 13. 3–219. 175

———, 1994. "Causality without Counterfactuals." *Philosophy of Science*, **61**(2): 297–312. 135

———, 1997. "Causality and Explanation: A Reply to Two Critiques." *Philosophy of Science*, **64**(3): 461–77. 135

———, 1998. *Causality and Explanation*. Oxford: Oxford University Press. 82, 110, 133, 134, 135

Salmon, W.C., Jeffrey, R.C., and Greeno, J.G., 1971. *Statistical Explanation and Statistical Relevance*. Pittsburgh, PA: University of Pittsburgh Press. 151, 175, 176

Sartwell, P.E., 1960. "'On the Methodology of Investigations of Etiologic Factors in Chronic Diseases' - Further Comments." *Journal of Chronic Diseases*, **11**(1): 61–3. 19

Savage, L., 2008. "Review of Swedish Cervical Cancer Screening Program Demonstrates Effectiveness and Room for Improvement." *Journal of the National Cancer Institute*, **100**(9): 603. 54

Scheffner, M., Werness, B.A., Huibregtse, J.M., Levine, A.J., and Howley, P.M., 1990. "The E6 Oncoprotein Encoded by Human Papillomavirus Types 16 and 18 Promotes the Degradation of p53." *Cell*, **63**(6): 1129–36. 73

Schiffman, M.H., Bauer, H.M., Hoover, R.N., Glass, A.G., Cadell, D.M., Rush, B.B., Scott, D.R., Sherman, M.E., Kurman, R.J., and Wacholder, S., 1993. "Epidemiologic Evidence Showing That Human Papillomavirus Infection Causes Most Cervical Intraepithelial Neoplasia." *Journal of the National Cancer Institute*, **85**(12): 958–64. 74

Schlehofer, J.R. and zur Hausen, H., 1982. "Induction of Mutations Within the Host Cell Genome by Partially Inactivated Herpes Simplex Virus Type 1." *Virology*, **122**(2): 471–5. 65

Schlumberger, H. and Lucké, B., 1949. "Serial Intraocular Transplantation of Frog Carcinoma for Fourteen Generations." *Cancer Research*, **9**(1): 52–60. 16, 127

Schmid, R. and Hammaker, L., 1961. "Hereditary Absence of Muscle Phosphorylase (McArdle's Syndrome)." *The New England Journal of Medicine*, **264**(5): 223–5. 98

Schmid, R. and Mahler, R., 1959. "Chronic Progressive Myopathy with Myoglobinuria: Demonstration of a Glycogenolytic Defect in the Muscle." *The Journal of Clinical Investigation*, **38**(11): 2044–58. 98

Schmid, R., Robbins, P., and Traut, R., 1959. "Glycogen Synthesis in Muscle Lacking Phosphorylase." *Proceedings of the National Academy of Sciences of the United States of America*, **45**(8): 1236–40. 98

Schneiderman, M.A. and Levin, D.L., 1973. "Parallels, Convergences, and Departures in Case-control Studies and Clinical Trials." *Cancer Research*, **33**(6): 1498–503. 64

de Schryver, A., Friberg, S., Klein, G., Henle, W., Henle, G., De-Thé, G., Clifford, P., and Ho, H.C., 1969. "Epstein-Barr Virus-associated Antibody Patterns in Carcinoma of the Post-nasal Space." *Clinical and Experimental Immunology*, **5**(5): 443–59. 28

Schwarz, E., Freese, U.K., Gissmann, L., Mayer, W., Roggenbuck, B., Stremlau, A., and zur Hausen, H., 1985. "Structure and Transcription of Human Papillomavirus Sequences in Cervical Carcinoma Cells." *Nature*, **314**(6006): 111–4. 73

Senior, K., 2008. "Global Aspirations for HPV Vaccination." *The Lancet Infectious Diseases*, **8**(3): 150. 75

Sever, J.L., 1973. "Herpesvirus and Cervical Cancer Studies in Experimental Animals." *Cancer Research*, **33**(6): 1509–10. 62

Shafritz, D.A., Shouval, D., Sherman, H.I., Hadziyannis, S.J., and Kew, M.C., 1981. "Integration of Hepatitis B Virus DNA into the Genome of Liver Cells in Chronic Liver Disease and Hepatocellular Carcinoma. Studies in Percutaneous Liver Biopsies and Post-mortem Tissue Specimens." *The New England Journal of Medicine*, **305**(18): 1067–73. 104

Sherlock, S., Fox, R.A., Niazi, S.P., and Scheuer, P.J., 1970. "Chronic Liver Disease and Primary Liver-cell Cancer with Hepatitis-associated (Australia) Antigen in Serum." *The Lancet*, **1**(7659): 1243–7. 102

Shope, R.E., 1932. "A Transmissible Tumor-like Condition in Rabbits." *Journal of Experimental Medicine*, **56**(6): 793–802. 16, 68

Shope, R.E. and Hurst, E.W., 1933. "Infectious Papillomatosis of Rabbits: with a Note on the Histopathology." *Journal of Experimental Medicine*, **58**(5): 607–24. 16, 68

Siegel, B.V., 1956. "Contaminant Filterable Agent Derived from a Human Wart." *The Journal of Investigative Dermatology*, **27**(6): 379–81. 67

———, 1960. "Electron Microscopic Studies of the Virus of Human Warts." *The Journal of Investigative Dermatology*, **35**(2): 91–3. 66

Siegel, B.V. and Novy, F.G., 1955. "Cultivation Studies on Wart Suspensions of Verruca Vulgaris and Condyloma Acuminatum." *The Journal of Investigative Dermatology*, **25**(4): 265–8. 67

Simons, P.J. and Ross, M.G., 1965. "The Isolation of Herpes Virus from Burkitt Tumours." *European Journal of Cancer*, **1**(2): 135–6. 48

Singh, P., 1955. "Proceedings of a Mulago Staff Clinical Meeting 29th October, 1955: Tumours of the Jaw." *The East African Medical Journal*, **33**(2): 70–1. 30, 160

Sixbey, J.W., Nedrud, J.G., Raab-Traub, N., Hanes, R.A., and Pagano, J.S., 1984. "Epstein-Barr Virus Replication in Oropharyngeal Epithelial Cells." *The New England Journal of Medicine*, **310**(19): 1225–30. 27

Skinner, G.R., 1976. "Transformation of Primary Hamster Embryo Fibroblasts by Type 2 Simplex Virus: Evidence for a 'Hit-and-Run' Mechanism." *British Journal of Experimental Pathology*, **57**(4): 361–76. 65

Skloot, R., 2010. *The Immortal Life of Henrietta Lacks*. London: Macmillan. 71

Spirtes, P., Glymour, C.N., and Scheines, R., 1993. *Causation, Prediction, and Search*. New York, NY: Springer-Verlag. 19

Stamler, J., Fields, C., and Andelman, S., 1967. "Epidemiology of Cancer of the Cervix. I. the Dimensions of the Problem: Mortality and Morbidity from Cancer of the Cervix." *American Journal of Public Health and the Nation's Health*, **57**(5): 791–803. 55, 58, 59

Stanley, N.F. and Keast, D., 1967. "A Reovirus-Specific Antigen in Murine Lymphoma 2731-L and in Cultured Burkitt's Lymphoma Cells." *The Australian Journal of Experimental Biology and Medical Science*, **45**(5): 517–25. 17, 47

Stanley, N.F., Walters, M.N., Leak, P.J., and Keast, D., 1966. "The Association of Murine Lymphoma with Reovirus Type 3 Infection." *Proceedings of the Society for Experimental Biology and Medicine*, **121**(1): 90–3. 17, 47

Stansfield, D., 1961. "Haematological Findings in African Children in Uganda with Malignant Lymphoma." *British Journal of Cancer*, **15**(1): 41–4. 36

Stavola, B.D., 1987. "Statistical Facts about Cancers on which Doctor Rigoni-Stern based his Contribution to the Surgeons' Subgroup of the IV Congress of the Italian Scientists on 23 September 1842. (translation)." *Statistics in Medicine*, **6**(8): 881–4. 56

Stebbing, J. and Bower, M., 2009. "Epstein-Barr Virus in Burkitt's Lymphoma: The Missing Link." *The Lancet Oncology*, **10**(4): 430. 26

Stell, P.M., 1977. "Epistaxis." *Clinical Otolaryngology*, **2**(3): 263–73. 23

Stern, E., Clark, V.A., and Coffelt, C.F., 1970. "Contraceptives and Dysplasia: Higher Rate for Pill Choosers." *Science*, **169**(3944): 497–8. 59

Stern, E. and Dixon, W.J., 1961. "Cancer of the Cervix–A Biometric Approach to Etiology." *Cancer*, **14**(1): 153–60. 56

Stewart, G.T., 1968. "Limitations of the Germ Theory of Disease." *Dental Digest*, **74**(11): 475. 18

Stewart, S.E., Lovelace, E., Whang, J.J., and Ngu, V.A., 1965. "Burkitt Tumor: Tissue Culture, Cytogenetic and Virus Studies." *Journal of the National Cancer Institute*, **34**(2): 319–27. 45, 46, 47

Stich, H.F., Hsu, T.C., and Rapp, F., 1964. "Viruses and Mammalian Chromosomes. I. Localization of Chromosome Aberrations After Infection with Herpes Simplex Virus." *Virology*, **22**(4): 439–45. 61

Stocks, P., 1955. "Cancer of the Uterine Cervix and Social Conditions." *British Journal of Cancer*, **9**(4): 487–94. 56

Stone, R.S., Shope, R.E., and Moore, D.H., 1959. "Electron Microscope Study of the Development of the Papilloma Virus in the Skin of the Rabbit." *The Journal of Experimental Medicine*, **110**(4): 543–6. 68

Strauss, M. and Shaw, E., 1949. "Crystalline Virus-Like Particles from Skin Papillomas Characterized by Intranuclear Inclusion Bodies." *Proceedings of the Society for Experimental Biology and Medicine*, **72**(1): 46–50. 66

Streissle, G. and Maramorosch, K., 1963. "Reovirus and Wound-Tumor Virus: Serological Cross Reactivity." *Science*, **140**(3570): 996–7. 47

Suchánková, A., Hirsch, I., Krcmár, M., and Vonka, V., 1984. "Determination of Herpes Simplex Virus Type-Specific Antibodies by Solid-Phase RIA on Helix Pomatia Lectin-Purified Antigens." *The Journal of Infectious Diseases*, **149**(6): 964–72. 63

Sumithran, E., 1977. "Rarity of Cancer of the Cervix in the Malaysian Orang Asli Despite the Presence of Known Risk Factors." *Cancer*, **39**(4): 1570–2. 56

Summers, J., Smolec, J.M., and Snyder, R., 1978a. "A Virus Similar to Human Hepatitis B Virus Associated with Hepatitis and Hepatoma in Woodchucks." *Proceedings of the National Academy of Sciences of the United States of America*, **75**(9): 4533–7. 104

Summers, J., O'Connell, A., Maupas, P., Goudeau, A., Coursaget, P., and Drucker, J., 1978b. "Hepatitis B Virus DNA in Primary Hepatocellular Carcinoma Tissue." *Journal of Medical Virology*, **2**(3): 207–14. 104

Suppes, P., 1970. *A Probabilistic Theory of Causality*. Amsterdam: North Holland Publishing Company. 82

Surgeon General, 1964. *Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service*. Washington: U.S. Dept. of Health, Education, and Welfare. 19, 166

Susser, M., 1991. "What is a Cause and How do we Know One? A Grammar for Pragmatic Epidemiology." *American Journal of Epidemiology*, **133**(7): 635–48. 19

Susser, M. and Rothman, K.J., 1988. *Causal Inference*. Chestnut Hill, MA: Epidemiology Resources. 19

Susser, M., 1995. "Invited Commentary on 'Causation and Disease: a Chronological Journey. The Thomas Parran Lecture'." *American Journal of Epidemiology*, **142**(11): 1125. 18

Sweet, B.H. and Hilleman, M.R., 1960. "The Vacuolating Virus, S.V. 40." *Proceedings of the Society for Experimental Biology and Medicine*, **105**(11): 420–7. 17

Syverton, J. and Berry, G., 1935. "Carcinoma in the Cottontail Rabbit Following Spontaneous Virus Papilloma (Shope)." *Proceedings of the Society for Experimental Biology and Medicine*, **33**: 399–400. 68

Tabery, J.G., 2004. "Synthesizing Activities and Interactions in the Concept of a Mechanism." *Philosophy of Science*, **71**(1): 1–15. 124

Tao, Q., Srivastava, G., Chan, A.C., Chung, L.P., Loke, S.L., and Ho, F.C., 1995. "Evidence for Lytic Infection by Epstein-Barr Virus in Mucosal Lymphocytes instead of Nasopharyngeal Epithelial Cells in Normal Individuals." *Journal of Medical Virology*, **45**(1): 71–7. 27

Temin, H.M., 1968. "Carcinogenesis by Avian Sarcoma Viruses." *Cancer Research*, **28**(9): 1835–8. 16

Temin, H.M. and Mizutani, S., 1970. "RNA-dependent DNA Polymerase in Virions of Rous Sarcoma Virus." *Nature*, **226**(5252): 1211–3. 16

Terris, M., Wilson, F., Smith, H., and Nelson, J., 1967. "Epidemiology of Cancer of the Cervix. V. The Relationship of Coitus to Carcinoma of the Cervix." *American Journal of Public Health and the Nation's Health*, **57**(5): 840–7. 56, 60

Thiry, L., Sprecher-Goldberger, S., Fassin, Y., Gould, I., Gompel, C., Pestiau, J., and Halleux, F.D., 1974. "Variations of Cytotoxic Antibodies to Cells with Herpes Simplex Virus Antigens in Women with Progressing or Regressing Cancerous Lesions of the Cervix." *American Journal of Epidemiology*, **100**(4): 251–61. 64

Thomas, D.B., 1973. "An Epidemiologic Study of Carcinoma *in situ* and Squamous Dysplasia of the Uterine Cervix." *American Journal of Epidemiology*, **98**(1): 10–28. 60

Thorley-Lawson, D.A. and Allday, M.J., 2008. "The Curious Case of the Tumour Virus: 50 Years of Burkitt's Lymphoma." *Nature Reviews Microbiology*, **6**(12): 913–24. 27

Tokatli, A., Coşkun, T., and Ozalp, I., 1998. "Isovaleric Acidemia. Clinical Presentation of 6 Cases." *The Turkish Journal of Pediatrics*, **40**(1): 111–9. 156

Tokuhata, G.K., 1967. "Epidemiology of Cancer of the Cervix. IV. Tobacco and Cancer of the Genitalia Among Married Women." *American Journal of Public Health and the Nation's Health*, **57**(5): 830–9. 59

Toms, A.P., Kasmai, B., Williams, S., and Wilson, P., 2006. "Building an Anonymized Catalogued Radiology Museum in PACS: A Feasibility Study." *British Journal of Radiology*, **79**(944): 666–71. 23

Towne, J., 1955. "Carcinoma of the Cervix in Nulliparous and Celibate Women." *American Journal of Obstetrics and Gynecology*, **69**(3): 606–13. 56

Trentin, J.J., Yabe, Y., and Taylor, G., 1962. "The Quest for Human Cancer Viruses: A New Approach to an Old Problem Reveals Cancer Induction in Hamsters by Human Adenovirus." *Science*, **137**(3533): 835–41. 17

Tulodziecki, D., 2007. "Breaking the Ties: Epistemic Significance, Bacilli, and Underdetermination." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, **38**(3): 627–41. 18

Turmezei, T., 2009. "The Surgical Sieve." *British Medical Journal*, **339**(b5384): 1450. 23

Ullman, E., 1923. "On the Aetiology of the Laryngeal Papilloma." *Acta Oto-Laryngologica*, **5**(2): 317–34. 66

Vandenbroucke, J.P., 2008. "Observational Research, Randomised Trials, and Two Views of Medical Science." *PLoS Medicine*, **5**: e67. 17

Vecchia, C.L., Franceschi, S., Decarli, A., Gentile, A., Fasoli, M., Pampallona, S., and Tognoni, G., 1984. "Dietary Vitamin A and the Risk of Invasive Cervical Cancer." *International Journal of Cancer*, **34**(3): 319–22. 59

Veldhuis, J.D., Garmey, J.C., Urban, R.J., Demers, L.M., and Aggarwal, B.B., 1991. "Ovarian Actions of Tumor Necrosis Factor-α (TNFα): Pleiotropic Effects of TNFα on Differentiated Functions of Untransformed Swine Granulosa Cells." *Endocrinology*, **129**(2): 641–8. 131

Villa, L.L., 2007. "Overview of the Clinical Development and Results of a Quadrivalent HPV (Types 6, 11, 16, 18) Vaccine." *International Journal of Infectious Diseases*, **11**(Suppl 2): S17–25. 75

de Villiers, E.M., Gissmann, L., and zur Hausen, H., 1981. "Molecular Cloning of Viral DNA from Human Genital Warts." *Journal of Virology*, **40**(3): 932–5. 71

Vogel, C.L., Mody, N., Anthony, P.P., and Barker, L., 1970. "Hepatitis-Associated Antigen in Ugandan Patients with Hepatocellular Carcinoma." *The Lancet*, **296**(7674): 621–4. 102

Vonka, V., 2000. "Causality in Medicine: The Case of Tumours and Viruses." *Philosophical Transactions of the Royal Society B: Biological Sciences*, **355**(1404): 1831–41. 63, 148

Vonka, V., Kanka, J., Hirsch, I., Závadová, H., Krcmár, M., Suchánková, A., Rezácová, D., Broucek, J., Press, M., and Domorázková, E., 1984a. "Prospective Study on the Relationship Between Cervical Neoplasia and Herpes Simplex Type-2 Virus. II. Herpes Simplex Type-2 Antibody Presence in Sera Taken at Enrollment." *International Journal of Cancer*, **33**(1): 61–6. 65, 171, 177

Vonka, V., Kanka, J., Jelínek, J., Subrt, I., Suchnek, A., Havránková, A., Váchal, M., Hirsch, I., Domorázková, E., and Závadová, H., 1984b. "Prospective Study on the Relationship Between Cervical Neoplasia and Herpes Simplex Type-2 Virus. I. Epidemiological Characteristics." *International Journal of Cancer*, **33**(1): 49–60. 59, 65, 171, 177

Vonka, V., Kanka, J., and Roth, Z., 1987. "Herpes Simplex Type 2 Virus and Cervical Neoplasia." *Advances in Cancer Research*, **48**: 149–91. 64

Vonka, V. and Hamsíková, E., 2007. "Vaccines Against Human Papillomaviruses–A Major Breakthrough in Cancer Prevention." *Central European Journal of Public Health*, **15**(4): 131–9. 75

Walker, L., Levine, H., and Jucker, M., 2006. "Koch's Postulates and Infectious Proteins." *Acta Neuropathologica*, **112**(1): 1–4. 18

Walter, E.L., Walker, D.L., and Cooper, G.A., 1965. "Localization of Specific Antigen in Human Warts." *Archives of Pathology*, **79**(4): 419–23. 67

Wang, C., Liu, P., and Liao, J.K., 2008. "Pleiotropic Effects of Statin Therapy: Molecular Mechanisms and Clinical Results." *Trends in Molecular Medicine*, **14**(1): 37–44. 131

Watts, S.L., Phelps, W.C., Ostrow, R.S., Zachow, K.R., and Faras, A.J., 1984. "Cellular Transformation by Human Papillomavirus DNA *in vitro*." *Science*, **225**(4662): 634–6. 73

Weber, E., 2009. "How Probabilistic Causation Can Account for the Use of Mechanistic Evidence." *International Studies in the Philosophy of Science*, **23**(3): 277–95. 86, 91, 92

Weed, D.L., 1995. "Causal and Preventive Inference." In P. Greenwald, B.S. Kramer, and D.L. Weed, editors, *Cancer Prevention and Control*, New York, NY: M Dekker, chapter 17. 285–302. 17

Weiss, L.M. and Movahed, L.A., 1989. "*In situ* Demonstration of Epstein-Barr Viral Genomes in Viral-associated B Cell Lymphoproliferations." *The American Journal of Pathology*, **134**(3): 651–9. 27

Wentz, W.B., Reagan, J.W., and Heggie, A.D., 1975. "Cervical Carcinogenesis with Herpes Simplex Virus, Type 2." *Obstetrics and Gynecology*, **46**(2): 117–21. 62

Wildy, P., 1973. "Antigens of Herpes Simplex Virus of Oral and Genital Origin." *Cancer Research*, **33**(6): 1465–8. 64

Wile, U.J. and Kingery, L.B., 1919. "The Etiology of Common Warts; Preliminary Report of an Experimental Study." *Journal of the American Medical Association*, **73**(13): 970–3. 66

Williamson, J., 2010. "Mechanistic Theories of Causality." Unpublished MS draft. 109, 110, 136

Wilson, J., 1937. "Genital Warts." *Journal of the Royal Army Medical Corps*, **68**: 227–38; 305–16. 66

Wimsatt, W.C., 1997. "Aggregativity: Reductive Heuristics for Finding Emergence." *Philosophy of Science*, **64**(Suppl II): S372–84. 111

Winkelstein, W., Shillitoe, E.J., Brand, R., and Johnson, K.K., 1984. "Further Comments on Cancer of the Uterine Cervix, Smoking, and Herpesvirus Infection." *American Journal of Epidemiology*, **119**(1): 1–8. 59

Witter, R.L., Burgoyne, G.H., and Solomon, J.J., 1969. "Evidence for a Herpesvirus as an Etiologic Agent of Marek's Disease." *Avian Diseases*, **13**(1): 171–84. 16

Woodall, J.P. and Haddow, A.J., 1962. "Isolation of an Agent from a Case of Burkitt's Lymphoma Syndrome." In A.J. Haddow and D. McCallum, editors, *East African Virus Research Institute Report 1961–2*, Nairobi Government Printer. 30–1. 48

Woodall, J.P., Williams, M.C., Simpson, D.I., and Haddow, A.J., 1965. "The Isolation in Mice of Strains of Herpes Virus from Burkitt Tumours." *European Journal of Cancer*, **1**(2): 137–40. 48

Woodward, J., 2000. "Explanation and Invariance in the Special Sciences." *British Journal for the Philosophy of Science*, **51**(2): 197–254. 115, 117, 118, 120

———, 2002. "What Is a Mechanism? A Counterfactual Account." *Philosophy of Science*, **69**(S3): S366–77. 117, 120

———, 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press. 117, 130, 131, 132

Wright, P.A. and Wynford-Thomas, D., 1991. "Koch's Postulates Revisited." *The Journal of Pathology*, **165**(4): 355. 18

Wynder, E., 1955. "Environmental Factors in Cervical Cancer; an Approach to Its Prevention." *British Medical Journal*, **1**(4916): 743–7. 59

———, 1966. "On Inferring Causality from Observed Associations." In F.J. Ingelfinger, A.S. Relman, and M. Finland, editors, *Controversy in Internal Medicine*, New York, NY: Sanders. 649–58. 19

Wynder, E., Cornfield, J., Schroff, P., and Doraiswami, K., 1954. "A Study of Environmental Factors in Carcinoma of the Cervix." *American Journal of Obstetrics and Gynecology*, **68**(4): 1016–47; discussion, 1048–52. 56, 58, 59

Yasumoto, S., Burkhardt, A.L., Doniger, J., and DiPaolo, J.A., 1986. "Human Papillomavirus Type 16 DNA-Induced Malignant Transformation of NIH 3T3 Cells." *Journal of Virology*, **57**(2): 572–7. 73

Yee, C., Krishnan-Hewlett, I., Baker, C.C., Schlegel, R., and Howley, P.M., 1985. "Presence and Expression of Human Papillomavirus Sequences in Human Cervical Carcinoma Cell Lines." *The American Journal of Pathology*, **119**(3): 361–6. 73

Yerushalmy, J., 1966. "On Inferring Causality from Observed Associations." In F.J. Ingelfinger, A.S. Relman, and M. Finland, editors, *Controversy in Internal Medicine*, New York, NY: Sanders. 659–68. 19

Yerushalmy, J. and Palmer, C.E., 1959. "On the Methodology of Investigations of Etiologic Factors in Chronic Diseases." *Journal of Chronic Diseases*, **10**(1): 27–40. 19

Zanetti, A.R., Damme, P.V., and Shouval, D., 2008. "The Global Impact of Vaccination Against Hepatitis B: A Historical Overview." *Vaccine*, **26**(49): 6266–73. 104

Ziegler, J.L., 1977. "Treatment Results of 54 American Patients with Burkitt's Lymphoma are Similar to the African Experience." *The New England Journal of Medicine*, **297**(2): 75–80. 27

———, 1981. "Burkitt's Lymphoma." *The New England Journal of Medicine*, **305**(13): 735–45. 27