EVIDENCE BASED MEDICINE, 'PLACEBOS' AND THE HOMEOPATHY CONTROVERSY

ANDREW JAMES TURNER, BA. MA.

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

JULY 2012

ABSTRACT

Homeopathic treatment has been available on the UK's National Health Service (NHS) since 1948. In recent years the continued provision of homeopathy through the NHS has been increasingly questioned as part of the ascendency of evidence-based medicine (EBM). Indeed, in 2009 the House of Common's Science and Technology committee commenced an 'Evidence Check' inquiry into Government policy supporting the NHS provision of homeopathic treatments. The controversy over whether homeopathic treatments 'really' work and whether they should be available through the NHS has generated much debate: at the heart of the controversy are questions about the nature of evidence in medicine, the validity of randomised trials and the nature and utility of 'placebo effects'. Critics of homeopathy put forward the simple argument that best available evidence shows homeopathic treatments to be equivalent to placebo, and therefore conclude that it should not be available through publically funded healthcare.

This thesis presents a critical examination of the concepts of EBM and 'placebos' and re-evaluates their role in the controversy around homeopathy. This thesis examines what kind of foundation the EBM philosophy of evidence provides for the arguments made in the controversy, and the role that 'placebos' play as both an evidential and normative standard.

There are two basic arguments: first, that the arguments justifying the EBM philosophy of evidence are fundamentally unclear, but also that the interpretation given to EBM, in debates about homeopathy, cannot be sustained. Second, that the concept of 'placebos' should be abandoned entirely: a framework is developed for talking about the effectiveness of treatments that removes much confusion about the epistemological and ethical standards that effective treatments should be held to. In addition to attempting to provide conceptual clarity to the controversy, the main conclusion is that the Science and Technology Committee have (on the basis of their own assumptions) understated their evidential arguments, by ignoring mechanistic evidence for whether homeopathic treatments are effective, and they have overstated their ethical arguments, they do not provide good reasons to remove provision of homeopathic treatment through the NHS.

ACKNOWLEDGEMENTS

The most thanks to my supervisors, Prof. Paul Martin and Prof. Cees van der Eijk for their ideas, guidance and support. Many thanks also to the Foundation for the Sociology of Health & Illness, for the studentship which has funded this PhD.

Thanks too, to everyone within the Institute for Science & Society, which has been a truly fantastic place to work. And thanks to Rachel Johnson, the source of many of the medical examples.

Note

An early version of Chapter 10 has been published as: Turner, A (2012) 'Placebos' and the Logic of Placebo Comparison, Biology & Philosophy, 27(3), 419-432. doi:10.1007/s10539-011-9289-8.

CONTENTS

PART ONE: THE HOMEOPATHY CONTROVERSY

CHAPTER 1

1. Introduction

Homeopathy is an alternative medicine currently available through the UK's National Health Service (NHS) as well as through private practice and on the high street. Homeopathy is unusual primarily because of the counter-intuitive means by which the treatments are prepared and prescribed (see §2.1). Questions about whether homeopathic treatments work and whether they ought to be available to patients have been asked throughout its 200 year history; however homeopathy has recently come under increased criticism, especially concerning its place on the NHS (see §2.2). Opponents of homeopathy argue that it is unscientific, that it doesn't work and that it shouldn't be available. Most notably in 2009/10, government policy about the use of homeopathy on the NHS was subject to an inquiry by the House of Common's Science & Technology Committee (STC). The STC concluded that homeopathic treatments were no better than placebos and that therefore should not be funded by the NHS, or even regulated as if they were a medicine by the Medicines and Healthcare products Regulatory Agency (MHRA).

The debate about whether homeopathic treatments work and what place they should have in modern healthcare is interesting because proponents and opponents of homeopathy fundamentally disagree about the ways that homeopathic treatments should be evaluated. Proponents argue that the concept of 'evidence-based medicine' (EBM), which is dominant in modern healthcare, presents a philosophy of evidence that is not properly equipped to deal with treatments which, like homeopathy, are premised on unconventional principles that challenge many of the assumptions of biomedical science. The debate is interesting for the further reason that the concept of 'placebo' is charged with a complex mix of evidential and normative force: it serves as the standard by which it is judged whether a treatment works and whether it can be ethically provided to patients.

The controversy about homeopathy raises issues about the nature of medical evidence. The controversy will be used as a way to examine of the concepts of EBM and placebos. The thesis is divided into four parts. The first three parts, deal with the homeopathy controversy, EBM and placebos respectively, and the fourth concluding part re-evaluates the controversy.

Part One introduces the homeopathy controversy. **Chapter 2** provides some background to what homeopathic treatment involves as well as to the contemporary criticisms of it. **Chapter 3** examines the criticisms in more detail and shows that the homeopathy controversy is composed of an evidential and a policy debate. It argues that there is a reasonably well-defined "Canonical Criticism" of homeopathy, the arguments of which have been most fully expressed in the STC report on homeopathy. **Chapter 4** serves to summarise the questions that will be addressed in the rest of the thesis. The questions concern the concepts of EBM and placebos: what kind of foundation do these concepts give to the arguments used in the controversy? Chapter 4 also serves to introduce Part Two.

Part Two examines EBM. **Chapter 5** argues that only very weak conclusions are drawn from the basic arguments put forward by proponents of EBM in the medical literature. The result is that the proper interpretation of EBM is unclear, which has given rise to (accusations of, at the very least) an interpretation that ranks different kinds of evidence as *categorically* better or worse than others. **Chapter 6** sets outs to examine whether more can be said about the interpretation of EBM through an electronic content-analysis and multidimensional scaling of a corpus of around 600 papers about EBM. It is argued that there is no single or stable interpretation of EBM in the medical literature. The literature is, and always has been, unclear about what the details of EBM amount to. Therefore **Chapter 7** aims to give some account of what the interpretation of EBM should be, drawing heavily on the recent work of philosophers of science John Worrall and Jeremy Howick. It is argued that the Categorical Interpretation of EBM is not defensible and that there are no a priori constraints on what kinds of methods can generate good evidence. **Chapter 8** summarises the results from Part Two and introduces Part Three.

Part Three examines placebos. **Chapter 9** reviews the research literature examining 'placebos' and 'placebo effects'. It is argued that this research speaks against the view that placebo effects are merely psychological phenomena or that they point to problems in the biomedical paradigm. On the contrary, placebo effects are the result of a wide range of factors, which act through specific physiological mechanisms: there is no single 'placebo effect'; and there are multiple mechanisms by which such effects are generated. **Chapter 10** takes this idea further and argues that 'placebos' and 'placebo effects' are concepts that should be abandoned, and that removing reference to 'placebos' forces one to be more specific about the

details of particular therapeutic contexts. **Chapter 11** draws out the implications of Chapters 9 and 10 more fully. Notably, borrowing a term from Adolf Grünbaum, the concept of a treatment's 'characteristic component' is introduced (that is the component of a treatment that makes it *that treatment* specifically). Crucially, it is argued that the efficacy of the characteristic component is important for ethical reasons. Furthermore, and contrary to the common idea that 'placebo treatments' are unethical, it is suggested that treatments with inefficacious characteristic components could be provided ethically. **Chapter 12** summarises the results from Part Three and introduces Part Four.

Part Four re-evaluates the homeopathy controversy in light of Parts Two and Three. **Chapter 13** argues that the STC report undervalues mechanistic evidence because it is based on a Categorical Interpretation of EBM. This is notable because of the strong claims that are often made in the basis of mechanistic evidence in the homeopathy controversy. It is also argued that when it comes to evaluating whether homeopathic treatment 'works' the key concern is with the efficacy of the characteristic component, however opponents of homeopathy who claim it does not 'work' must be seen as expressing an ethical objection to the reasons *why* it is effective. Chapter 13 also makes a tentative attempt to suggest circumstances in which the provision of homeopathic treatment would be ethically permissible. **Chapter 14** summarises briefly the overall conclusions of the thesis.

A more extensive, but still brief, overview can be gained from reading **Chapters 4, 8, 12** and **14** together.

CHAPTER 2

2. What is homeopathy?

In this chapter homeopathy and the controversy surrounding it are introduced. First a historical introduction is given and the key principles of homeopathy are described. Second the rise of criticism of homeopathy, over the last five years in described; in particular the political attention that homeopathy has received is described. Readers familiar with homeopathy can skip §2.1; readers familiar with the homeopathy controversy may prefer to skip Chapter 2 altogether.

2.1 Historical introduction and key principles

2.1.1 Hahnemann's homeopathy

The German physician Samuel Hahnemann (1755-1843) received a conventional 18th century medical education, but became disillusioned[1]. He was concerned that the medical knowledge of his contemporaries left them 'grop[ing] in the dark[2]' with regard to clinical practice. He could discern little evidence that, by practicing medicine, he was not a 'murderer or aggravator of the sufferings of [his patients][3]'. He was consequently critical of his contemporaries' use of aggressive medical techniques and speculative polypharmacy[4]. He explained:

> '[physicians] mixed more than one, indeed several different drugs in their so-called prescriptions and administered them in frequent large doses. Thus precious and fragile human life, so easily destroyed, was frequently placed in jeopardy at the hands of these perverted people, especially since bleedings, emetics,

[1] (Bivins, 2007; Coulter, 1975; Rothstein, 1992)
[2] Hahnemann quoted in (Coulter, 1975) p. 310
[3] Hahnemann quoted in (Coulter, 1975) p. 310
[4] See: (Coulter, 1975) pp. 319-351 for a comprehensive account of Hahnemann's criticisms of conventional medicine. See also (Rothstein, 1992) pp. 152-3

> purges, blistering plaster, fontanels[5], setons[6], caustics and cauterizations were also used[7]'

Hahnemann was, for example, outraged by the death of Leopold II of Austria in 1792; or more precisely, he was outraged at *Leopold's physician* for having performed four successive bleedings. Hahnemann writes:

> 'We ask, from a scientific point of view, according to what principles has anyone the right to order a second venesection when the first has failed to bring relief? As for a third, Heaven help us! But to draw blood a fourth time when the three previous attempts failed to alleviate! ...science pales before this[8]'.

Of course Hahnemann was not the only critic of conventional medicine at the start of the nineteenth century (as, for example, the reaction to the death of George Washington in 1799 illustrates[9]), many were aware of and reflected on the fact that medical practice was both speculative and brutal[10]. As a consequence of his discontent with medicine Hahnemann had begun his own experiments in the late eighteenth century, searching for a more empirically grounded and successful method for treating patients[11]. In his aims therefore Hahnemann can be compared to familiar names in the history of modern – 'evidence-based' – medicine[12]. The difference of course is that Samuel Hahnemann devised the homeopathic system of medicine; which on the modern view, is considered to be rather the opposite of evidence-based medicine.

---

[5] Making holes in the skull. A modern medical dictionary has the term 'fontanelle' denoting the soft areas on an infant's skull that have not yet fused together (Martin, 2007)

[6] Cord or cloth inserted into a wound for drainage, or to deliberately form a fistula (Martin, 2007)

[7] (Hahnemann, 1983) §54 (p. 50)

[8] (Coulter, 1975) p. 316

[9] (Cheatham, 2008; B. Cohen, 2005) See also (Bivins, 2007) p. 94

[10] See (Kaufman, 1971) ch. 1 for a strong statement of just how brutal. A very understated comment appears in (Nicholls, 1988) p. 17: '[19th century medicine was] a vigorous medical style which increased the risk of iatrogenic damage'.

[11] See: (Coulter, 1975) p. 311 and (Rothstein, 1992) p. 153

[12] Recall for instance Archie Cochrane's war-time experiences with tuberculosis, and his frustration that there was no evidence as to whether conventional treatments did more harm than good (Archie Cochrane Archive Catalogue: ALC/5 Health Services Research, 2008; Cochrane, 1945; Cochrane & Blythe, 1989)

Before introducing the criticisms of homeopathy further however, I now explain homeopathic treatments in a little more detail; first focusing on the key principles underlying the homeopathic system of medicine.

2.1.2 Four principles of homeopathic treatment

The U.S. National Library of Medicine gives the following description of the MeSH[13] term 'homeopathy':

> 'A system of therapeutics founded by Samuel Hahnemann (1755-1843), based on the Law of Similars where "like cures like". Diseases are treated by highly diluted substances that cause, in healthy persons, symptoms like those of the disease to be treated. The dilutions are repeated so many times that there is less than one molecule per dose and it is suggested that benefit is from the energetic life force of the original substance'[14]

This description gives a basic account of what homeopathy involves, by referring to a number of key ideas or principles. The enumeration and explanation of those principles are differently emphasised in other authors' explanations of homeopathy[15]. For the purposes of this discussion, it will be sufficient to characterise homeopathy in terms of four principles.

Initially however it should be noted that accounts of homeopathy typically present it as a uniform system that closely resembles Hahnemann's own accounts of homeopathy[16]. The description given below falls into that category, as does the quote above from the U.S. National Library of Medicine. In this type of account there is typically little mention of the variations that exist in the current practice of homeopathic treatment. As Antony Campbell has noted this way of presenting the

---

[13] **Me**dical **S**ubject **H**eading

[14] (U.S. National Library of Medicine, 2009) See similarly: the NIH's National Centre for Complementary and Alternative Medicine webpage about homeopathy <http://nccam.nih.gov/health/homeopathy/>

[15] Various authors pick out different numbers of principles that may characterise homeopathy. For example: 2 principles (Sense About Science, 2006; K. Smith, 2011; Vickers & Zollman, 1999); 3 principles (Milgrom, 2006a; U.S. National Library of Medicine, 2009; Van Wassenhoven & Ives, 2004); even 8 principles, (Guajardo & J. Wilson, 2005)

[16] See for example: (Bivins, 2007; Blackie, 1981; Clover, 1989; W. B. Jonas, Kaptchuk, & Linde, 2003; Leckridge, 1997; Vickers & Zollman, 1999)

homeopathic system gives 'the impression that the system is a kind of medical coelacanth, an anachronism that has survived from an earlier age[17]'. Never the less, the four ideas described below are elements in any homeopathic system meriting the name; though the fourth principle to a lesser extent, since it is not necessary[18]. Other elements of homeopathy, as Hahnemann conceived it, such as the psora view of disease[19] are not dealt with below. The purpose of the explanation below is to introduce the essential characteristics of homeopathic treatment in general, rather than consider in depth either the historical development of those characteristics or their interpretation in modern practice.

It is important to acknowledge however that the specific interpretation and application of those principles allows much diversity in practice. That is to say, there are a range of practices consistent with these general homeopathic principles[20]. It is worthwhile to give some indication of that variation. One area of variation is the selection of what dose, and what frequency of each dose, to treat patients with[21]. Furthermore there are three different ways in which homeopathic treatments can be delivered: classical, plural or complex, depending on what range of a patient's symptoms are taken into consideration and how many different treatments are used to cover them all[22]. Similarly the choice of treatment for a given set of symptoms is also subject to variation in practice[23]: different kinds of symptoms are differently emphasised – Again there is a three-way distinction: local, general and mental[24]. These differences are important in so far as they reflect varying configurations of

[17] (A. Campbell, 1984)
[18] The fourth principle describes the 'individualization' of Homeopathic treatments, but they do not have to be individualised – for example, they can be purchased over-the-counter. In this thesis it is individualised homeopathy, involving a consultation with a homeopathy, which is meant by the phrase 'homeopathic treatment', that is, treatment by a homeopath.
[19] (Hahnemann, 1983)
[20] (Leckridge, 2008) p. 129
[21] (Blackie, 1986; Borland, 1988; A. Campbell, 1984; Leckridge, 1997, 2008)
[22] Firstly there is the classical or uniciste method (These three approaches are mostly given their French names in the literature); this is the method whereby a homeopath will prescribe a single remedy that aims to cover as great a totality of the patient's symptoms as possible. The second and third methods involve giving multiple medicines. The pluraciste method involves multiple medicines that are tailored to different dimensions (local, general or mental) of a patient's illness, each with their own separate dosing regimes. Alternatively the complexiste method involves only a single tablet, but which contains multiple medicines. Homeopathy as practiced in France varies between all three of these methods: homeopathy in Britain typically takes the classical, uniciste, approach. See: (Leckridge, 2008) p. 136-7
[23] (Borland, 1988; A. Campbell, 1984; Leckridge, 2008)
[24] (Leckridge, 1997, 2008)

commitment to the general principles of homeopathy[25]. With these caveats regarding homeopathy *as it is practiced* in mind, the four ideas which I take to characterise the homeopathic system *in general* are as follows:

2.1.2.1 Similarity principle

Hahnemann's experiments with cinchona bark[26] are commonly cited as providing the foundation for the *similarity principle*[27]. The similarity principle is the fundamental principle of homeopathy[28]. It was well known to Hahnemann and his contemporaries that cinchona bark was effective for treating malaria[29,30]. In his translation of William Cullen's *Treatise of the Materia Medica*, where Cullen notes the effectiveness of cinchona bark in treating malaria, Hahnemann reports that after taking the bark himself[31] he experienced symptoms *characteristic of malaria*[32]. Hahnemann's idea was to see the two observations that (1) cinchona cures malaria and (2) cinchona causes malaria-like symptoms in healthy individuals, as an instance of a general principle that a cure for a disease will cause symptoms similar to the disease, in the healthy. This is his *similarity principle*. Indeed Hahnemann marshalled a range of historical and anecdotal evidence in favour of the similarity principle, including for example: a proposal from Hippocrates that hot drinks should be given to patients with a fever; the advice of many of his contemporaries that warming a burn aids recovery more than cooling it; and, the ability of Jenner's cow pox vaccination[33] to reduce the severity of small pox infection[34].

---

[25] (Leckridge, 2008)
[26] Sometimes called 'Puruvian Bark' or 'Jesuit's Bark'.
[27] See for example: (A. Campbell, 1984; Coulter, 1975; Danciger, 1987; Rothstein, 1992)
[28] It is the similarity principle which motivated Hahnemann to name his new medical system homeopathy – the connection is to the Greek 'homoios' (meaning 'like' or 'similar').
[29] (Bivins, 2007) p.89, (Coulter, 1975) pp. 360-2, (Rothstein, 1992) p. 153
[30] Which we now know is because of the quinine and other related alkaloids which it contains. (Druilhe, Brandicourt, Chongsuphajaisiddhi, & Berthe, 1988)
[31] In order to determine the effects it had on someone healthy: he did not have malaria himself.
[32] (Rothstein, 1992) p.153, (Coulter, 1975) p. 361
[33] The vaccination anology is sometimes used in contemporary discussions to add plausibility to the unintuitive ideas underlying homeopathy – See for example: (Fisher, 2010); see also the Implausibility Argument, below.
[34] See: (Coulter, 1975) pp. 371-5

This principle was formulated and developed by Hahnemann in a number of essays[35] and three substantial books[36]. His *Organon* (first published in 1810) provides a definitive statement of the similarity principle:

> '[homeopathic] therapy chooses from among all the remedies whose actions upon the healthy have been established that one which has the power and propensity to produce an artificial disease condition most similar to the natural one being treated[37,38,39].

2.1.2.2 Small doses

Hahnemann's treatments, like those of his conventional contemporaries, were often based on toxic ingredients; including for example: belladonna, arsenic, hemlock, opium and various animal venoms[40]. Administering these substances at conventional doses had considerable harmful effects on both the healthy and the ill. By the beginning of the 19th century Hahnemann had therefore begun using smaller doses in his experiments[41].

The way in which Hahnemann extrapolated the need for much smaller doses constitutes a second key idea in his homeopathic system. He began using a method of dilution which created *radically* low dosages[42]. First one part of a substance, solid

[35] (Hahnemann, 1852)
[36] (Hahnemann, 1805, 1904, 1983)
[37] (Hahnemann, 1983) §24 p. 26 - see also §70 pp. 68-70
[38] Since Hahnemann committed himself to the view that it was the '*totality* of symptoms of the natural disease' against which a treatment must be selected, he paved the way for the detailed and broad patient histories that are characteristic of homeopathic consultations today (Hahnemann, 1983) §70 pp. 68-70. See also, for example: (Owen, Leckridge, & Fisher, 2007) Ch. 3, (Nicolai, 2008) pp. 54-6 (and also Ch. 6).
[39] The similarity principle also suggests a method for determining the healing potential of particular substances. By giving those substances to healthy individuals, and observing their reaction, one can identify the symptom profile of the disease that substance will treat See: (Hahnemann, 1983) §106-8 & §110 (pp. 98-102). This is the logic behind the method of, what is called, homeopathic 'provings' or 'human pathogenic trials'. See: (Rothstein, 1992) pp. 154-5. See also: (W. B. Jonas et al., 2003) pp. 393-399, (Nicholls, 1988) p. 3 & 9.
[40] (Coulter, 1975) p. 400, (Rothstein, 1992) pp. 155-6.
[41] On account of the fact that conventional doses, as Coulter notes, the treatments 'caused severe aggravation' of many patient's symptoms (Coulter, 1975) p. 400 & 404.
[42] Described in almost every book or paper which mentions homeopathy. A nice description is (Nicholls, 1988) pp. 74-5. Described here is the process of diluting to one-hundredth of the

or liquid, was dissolved in ninety-nine parts solvent[43] and succussed (shaken) in the case of liquids, or triturated (mixed or rubbed) in the case of solids. This created the first dilution of 'one centesimal' or '1C'. The second dilution, '2C', was produced by taking one part of the 1C and adding ninety-nine parts solvent; again succussing or tritutating the mixture. This process of one-to-ninety-nine dilution followed by succussion or trituration is repeated up to the desired centesimal.

The concentration of a given centesimal is one hundredth of the concentration of the previous centesimal (hence the name); but substances cannot continue to be diluted forever. It is therefore commonly noted that beyond the 12C dilution there would be a less than 50% chance that the resulting dilution contained one molecule of the original substance[44]. Despite this fact, which would have been known to Hahnemann, he recommended dilutions around 30C[45]. There was much debate and variation in opinion about the appropriate level of dilution in homeopathic practice, both during and after Hahnemann's time. As Coulter notes: 'Hahnemann's more enthusiastic followers, moreover, continued to dilute drugs beyond the thirtieth centesimal[46]'. It should be noted that, whilst some level of dilution is characteristic of homeopathic treatments, there is not a particular level

---

concentration at each stage; but in modern practice different scales may be used, e.g. one-tenth.

[43] Water, alcohol or 'milk sugar' (lactose).

[44] Again this is described in almost every book or paper about homeopathy – see for example: (Nicholls, 1988) p. 75. In my opinion the calculation is never properly explained either. As an aside, there is frequent reference to Avogadro's constant in this regard; but it is not always clear what the relevance of this constant is to homeopathic dilutions. I assume that idea is that *one mole of a substance* cannot be divided into more than 6.022x10^23 parts, and therefore one part of any solution with a ratio of dilution exceeding 1:6.022x10^23 might not be *expected* to always contain any molecule of that substance. Clearly this puts a lower bound on the concentration of a substance in a fixed volume of solvent: namely, one molecule per fixed-volume (6.022x10^-23 mol/volume). The claim then is that homeopathic solutions' expected concentration is less than that lower bound, which is supposed to be entailed by the statement that they are diluted 'beyond Avogadro's constant' (sometimes called BRAN dilutions – Beyond the Reciprocal of Avogadro's Number), and is supposed therefore to entail the low probability that the final (12C and above) solutions contain any of the original substance. It seems to me however that one could reach this final conclusion without ever having to mention Avogadro's constant, but simply by considering the probability that at least one molecule of the n original molecules of the substance is carried through each successive dilution; since the truth of the claim clearly depends on how much one starts with. It seems an odd assumption to always start with one mole.

[45] (Hahnemann, 1983) §270 (p. 194). See also: (Coulter, 1975) p. 401, (Rothstein, 1992) p. 156, (Nicholls, 1988) p. 75.

[46] (Coulter, 1975) p. 402. To continue the quote: 'General von Korsakoff in Russia went as far as the 1500th centesimal, while Equerry Jenichen of Wismar carried dilution to the 2500th, 8000th, and even 16,000th,

which is definitive[47]. That is to say, it is not necessarily the case that homeopathic treatments are produced from dilutions beyond 12C.

2.1.2.3 Dynamization

It was Hahnemann's theory of *dynamization*[48] which made it possible for him to recommend, with coherence, dilutions close to and beyond 12C, and for other homeopaths to go to further stages of dilution. The idea is that succussing or triturating at each dilution-stage creates solutions which are dynamized. It is the dynamization of homeopathic dilutions that, according to Hahnemann, explains how they are able to remain effective, indeed become *more* effective, when the substance which is the basis of the treatment is at a level of dilution that rules-out any conventional pharmacological effect it might have. According to Hahnemann, dynamizing dilutions is a process which both reveals and refines a substance's therapeutic potential:

> 'homoeopathy develops the inner, spirit-like medicinal powers of crude substance to a degree hitherto unheard of and makes all of them exceedingly, even immeasurably, penetrating, active and effective… this process is called *dynamization* or *potentization*[49].

Given this theory of dynamization one can see why Hahnemann, and other homeopaths of his time as well as today, believe that characterising homeopathic treatments only in terms of their level of dilution, without referring to their dynamization, is to wholly *mis*characterise them. Hahnemann explains:

> 'Every day one still hears homoeopathic medicinal potencies referred to as *mere dilutions*, while they are in fact quite the opposite: trituration and succussion unlock the natural substances, uncover and reveal the hidden medicinal powers lying hidden in their soul[50]'.

[47] (House of Commons Science & Technology Committee, 2010) ev. 21
[48] Or *Potentization*
[49] (Hahnemann, 1983) §269 (p. 187-180) Original emphasis.
[50] (Hahnemann, 1983) §269 (p. 187-190) Original emphasis.

One might draw an analogy with the relationship between a cake and the cake-mixture. To argue that cake-mixture is a delicious complement to tea because cake is, is clearly to neglect that cake is *cooked* cake-mixture. And so, to argue that homeopathic treatments are not effective medicines because high dilutions are not, is to neglect that homeopathic treatments are *dynamized* high dilutions. Of course, this analogy ignores the major point of contention. While cooking clearly turns cake-mixture into a delicious complement to tea, it is controversial whether dynamization really does turn high dilutions into effective medicines.

2.1.2.4 Individualisation of Treatments

The fourth principle that will be taken to characterise homeopathic treatment is the individualisation of treatments. This principle is not necessary however: non-individualised homeopathy is not a contradiction. The fourth principle stems from the focus of homeopathic treatment on the individual patient, rather than conventional disease categories: treatments are tailored to the individual patient, not a disease. When homeopaths claim to treat the 'totality of symptoms', this must be understood in a much wider sense than in conventional medicine. It includes what would ordinarily be thought of as unrelated or idiosyncratic aspects of the patient's life. This is why consultations are typically more involved, for instance the *Desktop Guide to Complementary and Alternative Medicine* notes that:

> 'A first consultation may take 1½ hours or longer. Homeopaths take a thorough history and explore the patient's problems in much detail, with a view to finding the optimally matching homeopathic drug[51]'

Similarly, Bob Leckridge claims:

> 'The process of understanding a patient and being able to work out the most suitable homeopathic remedies for them involves us

[51] (E Ernst, Pittler, & Wider, 2006) p. 327

in approaching the patient in a more 'holistic' way and in trying to understand their individuality as part of the diagnostic process[52]'

Homeopathic history taking is therefore highly detailed. Leckridge describes how patients should be encouraged to describe and expand upon their symptoms in their own words; and how homeopath's should question and encourage patients to describe not only the specific, local, symptoms they experience, but also the more general aspects of their physiological and psychological well-being[53]. Of course however, how this process is conducted and how it leads to the prescription of a homeopathic treatment is subject to much variation in practice, as noted above. The key point however is that in applying the similarity principle and finding the appropriate treatment for a patient their 'symptoms' are given a much more extensive definition; which consequently means that the homeopathic consultation and history taking is more involved. This is what is meant by the notion that homeopathic treatment is individualised.

Having introduced the key principles that underlie homeopathic medicine, I now describe some details of the contemporary controversy.

2.2 The controversy

Two aspects of the contemporary controversy are highlighted below. First the rise of criticism in media, political and academic contexts (plus a brief historical note). Second more detail is given about the political aspect of the controversy.

2.2.1 Criticism of homeopathy

Debate about homeopathy in the nineteenth century was more prominent than today[54] and critical essays abound[55]. One early and often quoted[56] critical work is Oliver Wendell Holmes' essay *Homeopathy and its Kindred Delusions*, which attacks the similarity principle, small doses theory, and Hahnemann's 'psora'

[52] (Leckridge, 1997) pp. 7-8
[53] See in particular (Leckridge, 1997) pp. 36-40. See also, for similar accounts of the homeopathic consultation: (Skinner, 2001) pp. 13-8 and (Clover, 1989) Ch. 5
[54] (Google, 2010)
[55] See for example: (Holmes, 1842; Jameslindlibrary.org, 2011; Simpson, 1853)
[56] For example: (Baum & Edzard Ernst, 2009; Colquhoun, 2009a)

conception of disease[57] (see below for further explanation). What is particularly interesting is the prescience of Holmes' short 600 word preface to his book which, in summarising his own argument, anticipates many of the modern criticisms of homeopathy: or rather, illustrates that modern criticism of homeopathy contain few themes that are without historical precedent. Namely that:

(1) anecdotal evidence is insufficient to establish the efficacy of a treatment:

> 'statements, made by persons unacquainted with the fluctuations of disease and the fallacies of observation, are to be considered in general as of little or no value in establishing the truth of a medical doctrine or the utility of a method of practice'

(2) provision of homeopathy is unethical and indirectly harmful:

> 'Those kind friends who suggest to a person suffering from a tedious complaint, that he "Had better try Homoeopathy", are apt to enforce their suggestion by adding, that "at any rate it can do no harm". This may or may not be true as regards the individual. But it always does very great harm to the community to encourage ignorance, error, or deception in a profession which deals with the life and health of our fellow-creatures'

(3) the effects of homeopathy are equivalent to placebo:

> 'some patients may have been actually benefited through the influence exerted upon their imaginations… So long as the body is affected through the mind, no audacious device, even of the most manifestly dishonest character, can fail of producing occasional good to those who yield it an implicit or even a partial faith'

[57] (Holmes, 1842) - Holmes' comes to the conclusion that homeopathy is: 'a mingled mass of perverse ingenuity, of tinsel erudition, of imbecile credulity, and of artful misrepresentation, too often mingled in practice, if we may trust the authority of its founder, with heartless and shameless imposition'

Each of these claims will be repeated in Chapter 3, which discusses the arguments made about homeopathy in the last five years: it is worth reiterating that the quotes above were written in 1842. It should be noted therefore that while this thesis draws on the recent literature debating homeopathy it could, I suggest, have equally well and without needing substantive changes, drawn only on a pre-1900 literature debating homeopathy. While the volume and quality of the evidence base on which to assess homeopathy has increased since the nineteenth century, many of the arguments being made on that evidence-base have changed little. Even where the focus of debate is different, for instance concerning the place of homeopathy on the NHS, the premises of these arguments make historically familiar points concerning the acceptability of different kinds of evidence, and the nature and ethics of placebo treatments – the two themes of this thesis.

Turning to more recent instantiations of the homeopathy controversy: Homeopathic treatment has been available on the NHS since it was established in 1948[58]. In recent years the continued provision of homeopathy within the NHS has been increasingly questioned as part of the ascendency of evidence-based medicine; a concept which has become dominant in healthcare[59]. Whilst the popularity and criticism of homeopathy has been growing since the 1980's[60], both have become more prominent in the last five years.

In August 2005 the *Lancet* published a meta-analysis of clinical trials investigating the efficacy homoeopathic medicines, which found that 'when analyses were restricted to large trials of higher quality there was no convincing evidence that homeopathy was superior to placebo[61]'. The *Lancet* editorial accompanying this paper hoped the result would finally put 'an end to homeopathy', and lamented that the debate about whether homeopathy works had persisted 'despite 150 years of unfavourable findings[62]'. Unfortunately for the *Lancet*'s expectations, Shang et al's meta-analysis has failed to close the debate. Indeed almost the opposite is true. The

[58] (Nicholls, 1988)
[59] (Nadav Davidovitch & Filc, 2006)
[60] After having waned in popularity during most of the twentieth century: (Google, 2010; Nicholls, 1988)
[61] (Shang, Huwiler-Müntener, et al., 2005) p. 730
[62] (Editorial, 2005) p. 690

meta-analysis generated a number of direct responses[63], as well as being a focal point of the debate in many subsequent papers[64].

Following this, the profile of criticism of homeopathy has risen. In addition to the many academic papers criticising homeopathy and describing the growing pressure for restrictions on homeopathic practice in the UK, there has been renewed criticism of homeopathy in both the mass media[65] and political sphere[66], as well as in a number of prominent blogs[67].

After 2005 the media contribution to the debate about homeopathy increased: the lowest number of articles about homeopathy, published in the UK national press, in any one year post-2005 (32), is equal to the most number articles published in any one year pre-2005[68]. In May 2006, Michael Baum and other senior scientists signed an open letter, published in the *Times*, addressed to the directors of commissioning in all NHS Primary Care Trusts across the country, urging them to discontinue their use of homeopathy and their contracts with NHS homeopathic hospitals[69].

Since then numerous newspaper articles[70] have been published which assert that there is no evidence that homeopathy outperforms placebo[71], which question whether homeopathic treatments should therefore be available to the public[72], and which additionally question the extent to which the availability and popularity of

[63] See for example: (Bell, 2005; Boiron, 2011; Dantas, 2005; Fisher, 2006; Fisher, Berman, Davidson, Reilly, & T. D. B. Thompson, 2005; Frass et al., 2005; Gold et al., 2008; Kiene, Kienle, & von Schön-Angerer, 2005; Linde & W. B. Jonas, 2005; Lüdtke & Rutten, 2008; Oberbaum, S. R. Singer, & Frass, 2005; Pandolfi, 2011; D. Peters, 2005; Raoult, 2005; Rutten & Stolper, 2008, 2009; Shang, Huwiler-Müntener, et al., 2005; Shang, Jüni, Sterne, Huwiler-Müntener, & Egger, 2005; Skandhan, Amith, & Avni, 2005; Walach, W. B. Jonas, & Lewith, 2005; Paul Wilson, 2009)

[64] It has so far attracted 244 references in the Web of Knowledge database (as of Dec 1, 2011).

[65] For example, this selection from the Guardian: (Asthana & McKie, 2010; Boseley, 2010; Brooks, 2009; Butterworth, 2007; N. Cohen, 2007; Edzard Ernst, 2010a; Freeman, 2010; Goldacre, 2008a; Harris, 2011; Lipsett, 2008; Robbins, 2010a, 2010b, 2010c; Roberts, 2010; Rutherford, 2009; Sample, 2007, 2008, 2009a, 2009b, 2009c, 2010a, 2010b; Singh, 2009)

[66] (House of Commons Science & Technology Committee, 2010)

[67] (Colquhoun, 2011; Goldacre, 2011; Lewis, 2011)(Steven Novella, 2011)

[68] Search of LexisNexis Database for articles containing the words "homeopathy" or "homoeopathy" three or more times, Dec 2010.

[69] (Baum, 2006) See also: (Baum, 2004)

[70] Notably *the Guardian* has been a main arena for this debate. Something some of its readers have criticised it for, see: (Butterworth, 2007)

[71] (N. Cohen, 2007; Edzard Ernst, 2010a; Goldacre, 2007a; Harris, 2011; Lipsett, 2009; Sample, 2007, 2009a; Singh, 2009)

[72] (Asthana & McKie, 2010; Harris, 2011; Robbins, 2010c; Sample, 2008, 2009a, 2009c, 2010a)

homeopathy is a symptom of an increasingly 'irrational' society[73]. As well as greater media coverage these same critical themes have also been taken up in a number of recently published popular science books[74].

Furthermore a number of events initiated by campaigners who are critical of homeopathy have gained attention in the media. Most notably in 2006 and again in 2011, Newsnight, Sense About Science and the London School of Hygiene and Tropical Medicine conducted undercover investigations of private homeopaths and found them offering homeopathic prophylaxis for malaria as an alternative to conventional antimalarial drugs[75]. In 2010 the Merseyside Sceptics Society began a public campaign (the 10:23 campaign[76]) which protested against Boots' sale of homeopathic products on the high street, and held 'mass overdose' demonstrations around the country[77]. The British Medical Association (BMA) voted in support of a ban for homeopathy on the NHS at their 2010 annual conference[78]. Also in 2011 the Science Museum's 'Living Medical Traditions' exhibit generated much public criticism for allowing, what was perceived to be, an insufficiently-critical presence of homeopathy in the museum[79].

As will be described below, this increasing criticism of homeopathy has received its fullest expression in the House of Commons Science and Technology Committee's (STC's) 'Evidence-Check' report on homeopathy, published in February 2010. In late-2009 The STC performed an 'Evidence Check' of government policy relating to the NHS provision, and MHRA licensing, of homeopathic treatments.

In response proponents of homeopathy have claimed that much of this criticism displays a naive and overtly 'scientistic' style, especially the criticism in the media[80]. Also a number of EDMs have also been put forward in the House of Commons more recently. A parliamentary early day motion (EDM1240) was signed by over 200 MPs in March 2007, urging the government to ensure the continued place

[73] In 2007 Richard Dawkins devoted part two of his Channel 4 television series *Enemies of Reason* (titled 'The Irrational Health Service') to making exactly these points against homeopathy.
[74] (Goldacre, 2008b; R. Shapiro, 2008; Singh & Edzard Ernst, 2009)
[75] (Jones, 2006)(Jones & Ghosh, 2011)
[76] As a reference to the Avogadro constant (6.022x10^23): the number of atoms or molecules that constitutes one mole of a substance, which is supposed to be a reference to the dilution of homeopathic treatments.
[77] (10:23 Campaign, 2010; Society Of Homeopaths, 2010)
[78] (Deborah Cohen, 2010)
[79] (M. Baker & Davenport, 2011; Science Museum, 2011)
[80] (Milgrom, 2008, 2009a)

of homeopathy and other complementary and alternative medicines (CAMs) in the NHS[81]. These EDMs are generally supportive of homeopathy; they respond to the BMA's vote in 2010 (mentioned above) and to the STC's report[82]. Public criticism has affected Primary Care Trust commissioning, however. In two years since 2005 the number of prescription for homeopathic treatments halved and a quarter of trusts reduced funding for homeopathy[83]. In 2007 the West Kent Primary Care Trust conducted a review into its commissioning of homeopathic treatments, and decided to end provision of homeopathy; which after an independent review was finalised in 2009[84]. In 2010, the Greater Manchester Medicines Management Group advised the ten Primary Care Trusts in the region against funding homeopathic treatments[85].

The number of homeopathic hospitals in the UK has decreased in recent years as well. In 2008 the Tunbridge Wells Homeopathic Hospital was closed[86]. And it was decided in 2007 that the Royal London Homeopathic Hospital would be renamed the Royal London Hospital for Integrated Medicine; representing a change in focus that came into effect in September 2010[87].

Despite pressure to end provision of homeopathic treatment through the NHS however, the Department of Health maintain that the decision to commission homeopathic treatments on the NHS should remain with Primary Care Trusts: both to satisfy patient demand and comply with directives concerning the provision of homeopathy legislated by the European Commission[88] (explained below).

In broad terms the critics of homeopathy can be viewed as making the following argument: *homeopathic treatment works no better than placebo, therefore it should not be available through publically funded health care, nor regulated as if it were a medicine.* There are two debates here, one about the evidence itself and a further debate about its policy implications. The arguments used in these debates

[81] EDM 1240: <http://www.edms.org.uk/edms/2006-2007/1240.htm>
[82] See for example EDM284, EDM285, EDM286, EDM287, EDM387a2, put forward in June 2010 and EDM908, EDM1165 put forward in February 2010. However those from June 2010 have had significant amendments proposed; making them much less positive about homeopathy. See: <http://www.edms.org.uk/edms/2010-2011/284.htm> …/285.htm> …/286.htm> …/287.htm> …/387A2.htm> …/908.htm> …/1165.htm>
[83] (Praities, 2008a, 2008b)
[84] (West Kent Primary Care Trust, 2007a, 2009) see: <http://www.westkentpct.nhs.uk/Share_your_point_of_view/Archive/index.html>
[85] (Anekwe, 2010)
[86] It falls within West Kent PCT, see: (Praities, 2008a, 2008b)
[87] (Lewis, 2010)
[88] (Government Response to the Science and Technology Committee report 'Evidence Check 2: Homeopathy,' 2010). paras 47-8

will be described in more detail in Chapter 3. However more needs to be said about the STC report and the unusual regulatory schemes under which homeopathy falls.

2.2.2 Health policy and the House of Common's Science & Technology Committee

The House of Commons Science & Technology Select Committee[89] (STC), from late-2009 to early-2010, performed an 'Evidence Check[90]' of government policy relating to the NHS provision and MHRA licensing of homeopathic medicines. The preparation of the 'Evidence Check' report involved a call for submissions of written evidence to the STC, as well as two oral evidence sessions convened by the STC (involving two panels of 'experts' and key stakeholders[91])[92]. The committees conclusions and recommendations were published as 'Evidence Check 2: Homeopathy[93]' in February 2010. Throughout this thesis the STC report will be a key document for exploring the arguments put forward by critics of homeopathy (explained below). The key recommendations were that homeopathic treatment should not be funded through the NHS and it should not be regulated as a medicine by the MHRA. In their report the STC recommended that:

[89] The STC was reformed in October 2009 (after having been transformed into the Innovation Universities Science and Skills Committee in 2007) following calls for it to be re-established (Innovation Universities Science and Skills Committee Press Release, 2009).

[90] One particularly interesting aspect of the newly formed STC's work are its – so-called – 'Evidence Checks'; these are assessments of the coherence between government policy on a particular topic, and the evidence-base that ought to inform such policy. The Evidence Checks 'examine how the government uses evidence to formulate and review its policies' (House of Commons Science & Technology Committee, 2010) para 1.

[91] It is perhaps worth noting the members of the two panels:
(1) Paul Bennett, Professional Standards Director and Superintendent Pharmacist, Boots, Tracey Brown, Managing Director, Sense About Science, Dr Ben Goldacre, Doctor and Journalist, Professor Jayne Lawrence, Chief Scientific Adviser, Royal Pharmaceutical Society of Great Britain, and Robert Wilson, Chairman, British Association of Homeopathic Manufacturers Professor Edzard Ernst, Director, Complementary Medicine Group, Peninsula Medical School, Dr Peter Fisher, Director of Research, Royal London Homeopathic Hospital, Dr Robert Mathie, Research Development Adviser, British Homeopathic Association, and Dr James Thallon, Medical Director, NHS West Kent.
(2) Professor David Harper CBE, Director General, Health Improvement and Protection, and Chief Scientist, Department of Health, Mr Mike O'Brien QC, MP, Minister for Health Services, Department of Health, and Professor Kent Woods, Chief Executive, Medicines and Healthcare Products Regulatory Agency.

[92] All the written submissions and transcripts of the panel meetings are appended to the Evidence Check report itself (House of Commons Science & Technology Committee, 2010)

[93] (House of Commons Science & Technology Committee, 2010)

> 'to maintain patient trust, choice and safety, the Government should not endorse the use of placebo treatments, including homeopathy. Homeopathy should not be funded on the NHS and the MHRA [Medicines and Healthcare products Regulatory Agency] should stop licensing homeopathic products[94]'.

In July 2010 the Government response to the STC report was published[95]. The Department of Health's (DH) response maintained that homeopathy should continue to be provided through NHS and regulated by the MHRA; because it improves patient choice and, more fundamentally, because the Department of Health are not properly placed to intervene[96]. It is worthwhile giving some background to the reasons for the DH's view. Firstly, with regard to provision of homeopathic medicines the DH state:

> '[we do] not maintain a position on any complementary or alternative treatments, leaving decisions on their use *by* the National Health Service, *to* the National Health Service'[97]

Commissioning decisions, that is, decisions about which treatments are providable to patients, are made by Primary Care Trusts, not the DH. The DH further state that it would constitute a very 'unusual step' for the DH to interfere with Primary Care Trust's autonomy[98].

Secondly and more interestingly, with regard to the regulation of homeopathic medicines, the licensing regulation under which homeopathic medicines are categorised considers them to be harmless treatments for minor and self limiting illnesses (e.g. common cold). They have a place on the NHS in virtue of both their traditional and contemporary popularity[99]. The MHRA's regulation of

[94] (House of Commons Science & Technology Committee, 2010) para 157 – NOTE: the report is a 275 page document, consisting of the findings of the Science & Technology committee and an appendix consisting of transcripts of the STC's two oral evidence hearings, plus the written submissions received. In what follows citations of this document will refer to a paragraph number, for the committee's findings; and refer to the page number of the appendix for the transcripts and submitted written evidence: thus the first page of the appendix begins on ev. 1 – this is the numbering format in the document.
[95] (Department Of Health, 2010)
[96] (Department Of Health, 2010) paras 47-8
[97] (House of Commons Science & Technology Committee, 2010) Ev. 61 [original emphasis]
[98] (Department Of Health, 2010) para 48
[99] See especially: (Nicholls, 1988)

homeopathic medicines extends only so far as (1) requiring evidence for their safety and manufacturing quality, and (2) setting a limitation on the set of permissible medical claims that homeopathic medicines can make. In particular, the MHRA does not require homeopathic medicines to demonstrate their efficacy. The multiple regulatory schemes which provide this framework however can seem convoluted and are worth explaining.

The 1968 Medicines Act established the requirement for all medicines to be licensed; and required medicines to demonstrate evidence of their efficacy, in order to be granted a license[100]. When it was enacted in 1971 however, those medicines currently on the market – including homeopathic medicines on the market – were granted a 'License of Right[101]', meaning that they automatically received a license for their current indications. Hence they were permitted to make claims about being able to treat those indicated conditions, without having to provide evidence of efficacy.

In 1992 a 'Simplified Scheme' for the licensing of homeopathic medicines was introduced by the European Union Directive 92/37/EC[102], which allowed homeopathic medicines to be granted a license without providing evidence of efficacy, but which also did not permit them to make claims to treat specific conditions[103]. To resolve the subsequent state of affairs, where identical homeopathic medicines may or not be permitted to make medical claims depending on which scheme they were licensed under, the Government introduced the 'National Rules' Scheme in 2005 (the scope of which was provided for by article 16 of EU directive 2001/83/EC – that is to say, it does not *replace* the Simplified Scheme)[104]. The National Rules Scheme allows any homeopathic medicine (including those not formerly eligible for a license, even under the Simplified Scheme) to be licensed for minor and self limiting conditions (and thereby make claims about being able to treat those indicated conditions), without having to provide evidence of efficacy[105].

[100] (Medicines Act, 1968) Sec. 19 ss.1b
[101] (Medicines Act, 1968) Sec. 16 & 25
[102] This directive <http://goo.gl/NcxVc> is no longer in force, but was consolidated under Directive 2001/83/EC <http://goo.gl/Jf4kh> See: (European Parliament, 2001; European Parliament, 1992).
[103] See: (European Parliament, 2001: Articles 12-16)
[104] (Department Of Health, 2010) para 41, also: (European Parliament, 2001; MHRA Consultation Letter MLX 312, 2005)
[105] Furthermore the MHRA proposed to undertake a review of those homeopathic medicines with Licences of Right, especially where they were licensed for serious indications, in order to

Consequently, a homeopathic medicine (on the market in 2010) might be licensed under one of three different schemes and be indicated for particular conditions whilst being exempt from providing evidence of efficacy. It was this slightly convoluted and perhaps counter-intuitive position that prompted the STC to conduct their Evidence Check. However since this regulatory framework was largely determined by the EU directives noted above, the DH claim they are not in a position to prohibit the use of homeopathy on the NHS or by the NHS.

2.3 Summary

Homeopathy and the controversy surrounding it have been introduced. Firstly the principles that define homeopathy were described; which for the purposes of this thesis consisted of the similarity, small doses, dynamisation and individualisation principles. Collectively these define, in the most general terms, the unusual way in which homeopathic treatments are prescribed and produced

Second, the rise of contemporary criticism of homeopathy was briefly described, along with an account of the current regulation of homeopathic treatments. The key document in this regard is a report published by the House of Commons Science & Technology Committee, who evaluated the evidence-base for government policy relating to homeopathy.

The basic argument put forward is simply that: homeopathic treatment works no better than placebo therefore it should not be available through the NHS, nor regulated as if it were a medicine. In Chapter 3, the structure of this argument will be examined in more detail.

bring them under either the Simplified Scheme or National Rules Scheme (MHRA Consultation Letter MLX 312, 2005)

CHAPTER 3

3. What arguments are put forward in the contemporary controversy?

This chapter describes the arguments put forward in debates about homeopathy. It argues that the controversy is made up of both an evidential and policy debate. Importantly it is argued that opponents of homeopathy put forward a 'Canonical Criticism' of homeopathy.

The core debate in the controversy about homeopathy is evidential. It concerns whether or not homeopathy works and is expressed in the language of 'evidence-based medicine'. The kind of evidence needed to decide the matter is disputed however: opponents of homeopathy characterise the debate as being about whether homeopathic treatments are equivalent to placebo; whereas proponents occupy a range of positions that, in different ways, contest this characterisation.

The policy implications of the evidential debate are clearly important; as the discussion of the contemporary controversy in Chapter 2 shows, a policy debate, concerning the place of homeopathy on the NHS, for example, is also an arguably more prominent (but less fundamental) component of the controversy. As Edzard Ernst summarised, at the second of the STC's oral evidence sessions:

> 'If the NHS's commitment to evidence-based medicine is more than lip service then, surely, money has to be spent for treatments that are evidence-based, and homeopathy is not[106]'.

There are two different directions in which the controversy branches, in response to the evidential debate: first, if it is thought that homeopathy does work, then it is argued that the medical profession is confronted with a genuinely radical piece of new knowledge, which perhaps promises to widen the therapeutic and philosophical scope of the biomedical paradigm[107]. Second (and which constitutes by far the majority of discussion), if it is thought that homeopathy doesn't work, then this is taken to generate a series of policy questions concerning the availability,

[106] (House of Commons Science & Technology Committee, 2010) Ev. 46
[107] See for example: (Boiron, 2011)

funding and regulation of homeopathy (both as part of the NHS, and privately), which rely for answers on a series of ethical arguments surrounding the use of placebo treatments[108]. There are then two focal points in the controversy: one evidential, concerning whether or not homeopathy works; and one political, concerning the policy implications of views about the ethics of placebo treatments.

One interesting feature of the controversy is the asymmetry of the arguments put forward by proponents and opponents of homeopathy. I claim that what can broadly be called the opponents of homeopathy present a well-defined 'Canonical Criticism', which receives its fullest expression in the STC report. Proponents of homeopathy on the other hand occupy a range of positions which, in a variety of different (and possibly incompatible) ways, contest the Canonical Criticism. Owing to its stability across and within different literatures[109], what I call the Canonical Criticism will be the basis on which to introduce the debates in the controversy[110]. Firstly I discuss, in §3.1, the arguments that make up the evidential debate; then in §3.2, the policy debate.

3.1 The evidential debate

3.1.1 The Canonical Criticism

The Canonical Criticism presents an account of both how to determine whether homeopathy works, and an evaluation of whether it does in fact work. That is to say, it presents an account both of what counts as evidence and what the evidence that counts tells us, as follows:

[108] (Asthana & McKie, 2010; Baum, 2006; Deborah Cohen, 2010; Edzard Ernst, 2008; Garattini & Bertelé, 2009; Harris, 2011; Hay, 2008; Hunter, 2002; NHS, 2010; O'Dowd, 2009; Robbins, 2010b, 2010c; Sample, 2008, 2010a; K. Smith, 2011; West Kent Primary Care Trust, 2007b; C. White, 2010; Yu-Hin Ng, 2011)

[109] Academic, media, political & internet.

[110] To be clear about the terms being used: the STC exemplifies the Canonical Criticism; but the Canonical Criticism is wider than simply the STC's arguments. The point is that much of the critical literature puts forward the same 'canonical' set of arguments; of the critical literature, the STC report is clearest and most explicit. As will be shown, there is a significant point of divergence between the Canonical Criticism and the STC concerning the role that mechanistic evidence is supposed to play.

(1) Evidence-based medicine provides the framework for assessing whether homeopathy works. It is a question of efficacy: do homeopathic treatments outperform placebo in randomised trials.

(2) The best available evidence (from randomised trials, or better, meta-analyses of such trials) shows that homeopathic treatment is equivalent to placebo.

(3) The homeopathy=placebo hypothesis is supported by mechanistic evidence which shows that it is implausible to expect homeopathic treatments to be efficacious.

These three points, and the way they are contested, will be explained in turn below:

3.1.2 What counts as evidence?

The justification for the Canonical Criticism's view about what counts as evidence derives from a philosophy of evidence called 'evidence-based medicine' (EBM). The core argument of the EBM view is that, as a guide to efficacy, one should trust the results of controlled clinical research over expert opinion or mechanistic theory[111]. Controlled clinical research is best placed to distinguish between the efficacy of a treatment and the contribution of other confounding factors.

The reason for the focus on efficacy is that medical treatments can be effective without being efficacious. Therapeutic effects can occur independently of whether the medicine itself was causing those effects[112]. This is the much referenced distinction between efficacy and effectiveness. The STC devote a section of their report to outlining precisely this distinction between efficacy and effectiveness. In their view, the key to the distinction is whether a treatment is a 'placebo' treatment or not. They state:

> 'If homeopathy was better than a placebo treatment, one would expect tests of efficacy to show that it is efficacious; and "real world" tests of effectiveness to show that it may or may not be effective. If homeopathy was a placebo

[111] For example: (Evidence Based Medicine Working Group, 1992)
[112] See also, for example: (I. Evans, Thornton, Chalmers, & Glasziou, 2011)

> treatment, it would fail tests of efficacy, but with tests of effectiveness it would appear to be effective for some conditions and some patients, but not for others[113]'

Proponents and opponents of homeopathy disagree about what is causing the appearance of therapeutic benefit: the homeopathic treatments themselves (efficacy), or, as Edzard Ernst suggests in his written submission to the STC Evidence Check report: '[patients may improve] because of placebo-effects, regression towards the mean, concomitant treatments and many other confounders[114]' (effectiveness). The point the STC and the Canonical Criticism emphasise, is that effectiveness is insufficient to infer efficacy.

The Canonical Criticism holds that what matters, when one asks whether homeopathy works, is that it is effective because it is efficacious. The reasoning behind this is that it would be wrong to claim homeopathy works, if one knows that homeopathic medicines are a *redundant component* in an explanation of what makes the treatment effective. Again, the STC make this clear:

> 'We have set out the issue of efficacy and effectiveness at some length to illustrate that a non-efficacious medicine might, in some situations, be effective (patients feel better) because of the placebo effect. *That is why* we put more weight on evidence of efficacy than of effectiveness[115]'

Thus the key question for opponents is *why* patients benefit from homeopathy. In a recent interview in the *British Medical Journal*, Edzard Ernst (perhaps the most prominent critic of homeopathy) makes precisely this point:

> 'Today he [Ernst] still accepts that homoeopathic treatments work—"the question is: why?" He says he now has a conclusive

[113] (House of Commons Science & Technology Committee, 2010) para. 28
[114] (House of Commons Science & Technology Committee, 2010) ev. 27
[115] (House of Commons Science & Technology Committee, 2010) para. 39 (My emphasis)

answer: "It works because of a very long empathetic consultation. It's a non-specific effect"[116]'

The STC put more weight on evidence of efficacy because it is not acceptable for treatments to work through placebo effects[117], as the emphasised section of the quotation from the STC report above illustrates. The Canonical Criticism holds that EBM provides the epistemological and methodological resources to best answer the question of whether homeopathic treatment works[118]. As the STC again state:

'If homeopathic products – or any medicinal product – are more than placebos, and all other elements of the "holistic" care package are the same (controlled), it should be possible to see differential results between the test substance and the placebo[119]'

The important point here is that the question of whether homeopathy works is framed as a question about the efficacy of homeopathic treatment. Whilst homeopathic treatment may be effective for a range of reasons, the Canonical Criticism holds that the only legitimate sense in which it can be said to 'work' is if its effectiveness is a direct consequence of its efficacy. The Canonical Criticism claims that the efficacy of homeopathy is demonstrated by the ability of homeopathic treatments to outperform placebo in randomised trials. This is the justification for (1), above.

3.1.3 Contesting what counts as evidence

Proponents of homeopathy contest the Canonical Criticism's framing of the evidential debate in a variety of ways. Below two of the main challenges are noted: first, that the interpretation of EBM in the Canonical Criticism is naïve and

[116] (D. Cohen, 2011)
[117] This will be returned to in Part Three – Placebo is a normative standard too.
[118] Some examples of where this view can be found: (Baum & Edzard Ernst, 2009; A. D. Boer & Porsius, 1997; Butterworth, 2007; Edzard Ernst, 2009a; Goldacre, 2007a; Hoffer, 2003; Renckens, 2002) See also the following reflections on the role of EBM: (Barry, 2006; Hansen & Kappel, 2010)
[119] (House of Commons Science & Technology Committee, 2010) para 22. The same point is frequently made elsewhere, see for example: (Goldacre, 2007a; Oberbaum, Vithoulkas, van Haselen, & S. R. Singer, 2003; Pandolfi, 2010)

unsophisticated. Second, that homeopathic treatment is a 'complex intervention[120]' and therefore not suited to being evaluated in placebo controlled trials. Plus a third point, made by those proponents of homeopathy who do think that randomised trials are appropriate.

Lionel Milgrom presses the first challenge most consistently[121], but it is also made by other proponents of homeopathy[122]. It is argued that the interpretation of EBM in the Canonical Criticism is 'scientistic'[123], and that focusing only on the results of placebo controlled trials fails to not provide the range of evidence needed to evaluate whether homeopathy works. That is to say, proponents of homeopathy argue that the question of whether homeopathy works cannot be sufficiently answered by evidence from randomised trials, because other evidence is also necessary. For example, Milgrom states that:

> 'EBM as currently practiced, now concentrates solely on the "gold-standard" double-blind randomized-controlled trial (DBRCT) and meta-analyses as the only acceptable scientific evidence for a therapy or procedure… [which results] in a downgrading and/or ignoring of other valid forms of evidence[124]'

> 'The RCT has (some have said) brutally displaced other forms of evidence-gathering, and is now regarded as the only proper way of gauging the efficacy of any drug or clinical procedure[125]'

> '[Opponents of homeopathy put forward] an evidence 'mono-culture', where the primacy of an 'ideal' scientifically-determined efficacy would subsume other no less important forms of

[120] For an introduction to the evaluation of complex interventions, see for example: (Medical Research Council, 2000) (P. Craig, P. Dieppe, et al., 2008)
[121] (Milgrom, 2005, 2008, 2009a, 2009b, 2010a) See also Milgrom's submission to the STC Report: (House of Commons Science & Technology Committee, 2010) Ev. 94-100
[122] See for example: (House of Commons Science & Technology Committee, 2010) Memorandum submitted by Dr Sara Eames Ev. 135; Memorandum submitted by the Society of Homeopaths Ev. 139; Memorandum submitted by the Alliance of Registered Homeopaths Ev. 152. See also: (Bell, 2005; Chatfield, 2008; W. B. Jonas, 2001; Walach, 2001)
[123] (House of Commons Science & Technology Committee, 2010) Ev. 95
[124] (House of Commons Science & Technology Committee, 2010) Ev. 94-5
[125] (Milgrom, 2009b) p. 205

> evidence, to the possible detriment of patient and clinician concerns[126]'

The problem identified here is that randomised trials have, according to proponents of homeopathy, been reified in the Canonical Criticism. Milgrom also argues that the Canonical Criticism's reification of randomised trials is not consistent with the EBM philosophy of evidence, as it *should* be interpreted[127]. He quotes favourably a criticism of EBM made by Michael Rawlins, namely that:

> 'RCTs, long regarded as the 'gold standard' of evidence, have been put on an undeserved pedestal. Their appearance at the top of hierarchies of evidence is inappropriate; and hierarchies are illusory tools for assessing evidence. They should be replaced by a diversity of approaches that involve analysing the totality of the evidence base[128,129]

Although Milgrom argues that the Canonical Criticism overvalues evidence from randomised trials, other authors have never the less claimed that such evidence is an important measure of *credibility* in the medical profession; and therefore ought to be the focus of research efforts into alternative medicines. As Oberbaum, Singer, & Frass argue:

> 'because RCTs remain the central pillar of evidence-based medicine, we believe that we at this stage concentrate our resources on this study design... even exceptional results obtained in unblinded, uncontrolled or observational studies will not carry home the point that homeopathy is indeed effective[130]'

[126] (Milgrom, 2010b) p. 84
[127] (Milgrom, 2005, 2008, 2009a, 2009b, 2010a)(House of Commons Science & Technology Committee, 2010) Ev. 94-100
[128] (M. D. Rawlins, 2008) quoted in: (Milgrom, 2009a, 2009b, 2010b)
[129] In fact, other opponents of homeopathy also make the same criticism of EBM; see for example (Steven Novella, 2011) <http://www.sciencebasedmedicine.org/index.php/homeopathy-and-evidence-based-medicine-back-to-the-future-part-v/>
[130] (Oberbaum et al., 2005) p. 304.

Indeed such is the authority of randomised trials that there are both proponents and opponents who have called for 'decisive' trials to be performed to clinch and close the controversy[131].

A second challenge that is made by proponents of homeopathy concerns the problems with using placebo controlled trials as a method for determining whether homeopathic treatments work[132]. It is claimed that there is something inherently different about how homeopathic treatments work, as compared to conventional medicines. For example Iris Bell claims:

> 'the very nature of homeopathy… is inherently non-specific… that they are not allopathic drugs [that is, conventional drugs], is consistent with the claims of homeopathic clinicians and the conceptual principles of the field[133]'.

A common theme in this kind of challenge is that some notion of complexity inherent in homeopathic treatments prevents placebo controlled trials (PCTs) providing good evidence for whether they work[134]. It is argued that this complexity justifies the view that homeopathic treatments are not suited to being evaluated in PCTs; because the distinction between the active and inactive elements of homeopathic treatment is blurred; with no meaningful way to pull them apart. The notion of a 'complex intervention' is acknowledged within biomedicine also. In 2000 the Medical Research Council published a *Framework for the Development and Evaluation of RCTs for Complex Interventions to Improve Health*[135], which was updated in 2008[136]. In those documents complex interventions are characterised by the fact that they 'contain several interacting components[137]', such that it is difficult to specify what the 'active ingredient[138]' truly is. This is precisely the claim that is made on behalf of homeopathic treatments, by its proponents. Examples of complex

[131] (Baum & Edzard Ernst, 2009; Oberbaum et al., 2005)
[132] (Fisher, 2009; T. D. B. Thompson & Weiss, 2006; Weatherley-Jones, E. A. Thompson, & K. J. Thomas, 2004)
[133] (Bell, 2005) p. 765
[134] (Fisher, 2009; Milgrom, 2005, 2006b, 2009a)
[135] (Medical Research Council, 2000) (M. Campbell et al., 2000)
[136] (N. C. Campbell et al., 2007; P. Craig, P. Dieppe, et al., 2008; Peter Craig, Paul Dieppe, et al., 2008)
[137] (Peter Craig, Paul Dieppe, et al., 2008) p. 7 – but see also: (M. Campbell et al., 2000; N. C. Campbell et al., 2007; P. Craig, P. Dieppe, et al., 2008; Medical Research Council, 2000)
[138] See (Medical Research Council, 2000) p. 1

interventions include medical treatments such as physiotherapy[139], and surgical procedures[140] as well as social interventions such as the Sure Start program[141], or stroke rehabilitation units[142]. As these documents show, the mere fact that a treatment might be a complex intervention does not rule out, in principle, the investigation of their effectiveness in rigorous experiments. What they do highlight however is the methodological sophistication often associated with doing so. Indeed the authors of the updated document caution one to: 'Beware of 'blanket' statements about what designs are suitable for what kind of intervention[143]'.

In these terms then, opponents and proponents of homeopathy can be seen to be contesting whether homeopathic treatment is a complex intervention. The Canonical Criticism holds the view that it is not; contrary to this there are a number of reasons put forward by proponents of homeopathy, for why homeopathic treatment should in fact be considered a complex intervention[144].

First, there are arguments based around the individualisation of homeopathic treatment (the fourth principle noted in §2.1). The claim is that, because homeopathic treatment treats the totality of a patient's symptoms with a medicine specific to that individual patient, this presents a barrier to averaging across patients receiving different treatments. Furthermore it is also claimed that outcome measures fail to capture the holistic nature of the improvement from homeopathic treatment[145]. For example, there are many statements in the literature similar to the following:

> 'homeopathy, as practised in the clinic, is singularly unsuited to the stipulations of the modern scientific method. Whereas medical research typically examines a single intervention for a given ailment, individualisation is the homeopathic dictum. The formal disease classification used in conventional medicine and research is largely irrelevant. The outcomes measured in

[139] (Medical Research Council, 2000) p. 1
[140] (Peter Craig, Paul Dieppe, et al., 2008) p. 20
[141] (P. Craig, P. Dieppe, et al., 2008)
[142] REF
[143] (Peter Craig, Paul Dieppe, et al., 2008) p. 10
[144] See notably: (Weatherley-Jones et al., 2004)(T. D. B. Thompson & Weiss, 2006)
[145] (Weatherley-Jones et al., 2004)(Bell, 2005)

> medicinal research are necessarily one-dimensional, whereas homeopathic outcomes are multifarious[146]'

Second there are arguments based around interactions between the 'active' and 'inactive' ingredients in homeopathic treatments. The claim is that these interactions made it difficult, if not impossible, for a clear distinction to be drawn between them. Consequently it is argued that the efficacy of homeopathic treatment is not easily, or cannot be, captured in PCTs[147]. Weatherley-Jones et al put the point most explicitly:

> 'The interaction of the non-specific effects of the consultation with the specific effects of the medicine appears to challenge the double-blind placebo-controlled RCT as a meaningful test of individualised homeopathy[148]'

Again:

> 'The fundamental concept of the placebo-controlled RCT as a method of estimating the size of the specific effect of treatment is thus inappropriate in therapies where there is potentially an interaction between the non-specific and specific effects of treatment[149]'

A number of proponents of homeopathy have offered explanations of these kinds of interactions, which draw on the notion of entanglement in quantum physics[150]; other authors draw on literature discussing complex systems[151]. The important point to note here is that the claim that homeopathic treatments are complex interventions is supposed to entail that the different components of

[146] (Oberbaum et al., 2005) p. 303
[147] For example: (Weatherley-Jones et al., 2004)(Fisher, 2009) (House of Commons Science & Technology Committee, 2010) ev. 167
[148] (Weatherley-Jones et al., 2004) p. 188, see also: (Milgrom, 2006a) p. 213
[149] (Weatherley-Jones et al., 2004) p. 188
[150] Notably: (Milgrom, 2005, 2006a, 2006b, 2009c; Walach, 2003, 2005; Weingärtner, 2007)
[151] (Bell & Koithan, 2006)

homeopathic treatment cannot be meaningfully pulled apart and investigated. It is this idea that is used to argue that PCTs of homeopathic treatments are methodologically inappropriate: precisely because PCTs separate and single-out one particular component of a treatment, controlling for the rest (see the quote from the STC above, in §3.1.2).

Importantly, an explanation of the inappropriateness of placebo controlled trials of homeopathic treatment also serves as an explanation of negative results from such trials[152]. If homeopathic treatments produce negative results in PCTs, that is unsurprising given that such trials are methodologically questionable[153]. Consequently these arguments about the appropriateness of PCTs have a dual purpose, firstly in debates about what should count as evidence, as well as in debates about what the evidence base for homeopathy is[154].

In opposition to this opponents argue that homeopathic treatments – like all candidate medical treatments, whatever their nature – fall under the logic of EBM and are amenable to properly designed controlled clinical research. If homeopathic treatments are atypical, then this is taken to be at most an issue for trial-designers, not an in principle objection to efficacy testing[155]. Thus for example we find statements such as the following:

> 'alternative practitioners have held the opinion that the super-individually adjusted approach of their patients precluded the possibility of randomized trials. This argument is, in these days of evidence-based medicine, no longer acceptable[156]'

The third point to note, in addition to these two challenges (that EBM is naïve and that randomised trials are not appropriate) is that, interestingly, not all proponents of homeopathy deny that PCTs are an appropriate test of whether homeopathic treatments are efficacious. Other authors have argued that homeopathy can legitimately be assessed in PCTs; for example Peter Fisher (Clinical Director of – what was formerly – the Royal London Homoeopathic Hospital) has

[152] (House of Commons Science & Technology Committee, 2010) ev. 169
[153] (Weatherley-Jones et al., 2004)
[154] See for example: (Keshet, 2009) p. 134.
[155] (Overall & Dunham, 2009) p. 148.
[156] (Renckens, 2002) p. 528.

claimed: 'Randomised placebo-controlled trials are, in principle, capable of demonstrating such effects for homeopathic treatments[157]'; and the British Homeopathic Association and Faculty of Homeopaths joint written submission to the STC also emphasised that their assessment of the evidence-base for the efficacy of homeopathic treatments 'focuses primarily on systematic reviews of published RCTs[158]'. Similarly, just over one third of submissions to the STC Evidence Check that were supportive of homeopathy made positive claims about the results of homeopathic treatments in clinical trials[159] – presumably thereby endorsing the view that homeopathic treatments are in principle testable.

In general terms therefore, the issue of how to test homeopathic treatments is largely an argument about the influence of EBM on the structure of the evidential debate. Is EBM interpreted properly in the Canonical Criticism? Are placebo controlled trials appropriate?

3.1.4 Meta-analyses of homeopathic treatments

The Canonical Criticism holds that evidence from randomised trials shows that homeopathic treatments are placebos[160]. Claims of this sort typically rely on results from a number of large-scale meta-analyses that have been performed in the last twenty years[161], as well as other smaller more specific reviews and analyses[162]. Opponents of homeopathy typically state their assessments of the clinical research evidence in the following terms:

[157] (House of Commons Science & Technology Committee, 2010) ev. 21
[158] (House of Commons Science & Technology Committee, 2010) ev. 37
[159] 12/30
[160] See for example: (A. D. Boer & Porsius, 1997; Editorial, 2005; Edzard Ernst, 2007, 2009a, 2009b, 2011a; Goldacre, 2007b; Pandolfi, 2010; Sehon & D. Stanley, 2010) and further examples in the mass media: (BBC Radio 4, 2005; N. Cohen, 2007; Goldacre, 2007a; Sample, 2010b; Sense About Science, 2006) and political sphere: (House of Commons Science & Technology Committee, 2010) Ev. 118, Ev. 27;
[161] For example: (Boissel, Cucherat, M. Haugh, & Gauthier, 1996; Cucherat, Margaret Haugh, Gooch, & Boissel, 2000; Feder & T. Katz, 2002; Kleijnen, Knipschild, & ter Riet, 1991; Linde & Melchart, 1998; Shang, Huwiler-Müntener, et al., 2005)
[162] For example: (Altunç, Pittler, & Edzard Ernst, 2007; Edzard Ernst, 1998, 1999; Long & Edzard Ernst, 2001; McCarney, Lasserson, Linde, & Brinkhaus, 2004; McCarney, Warner, Fisher, & van Haselen, 2003; Milazzo, N. Russell, & Edzard Ernst, 2006; Pilkington, 2005; Pilkington, Kirkwood, Rampes, Fisher, & J. Richardson, 2006)

> ‘the ~150 published trials collectively fail to indicate clinical effectiveness[163]’;
>
> ‘dozens of such reviews [of homeopathy] are available today. The vast majority of those that are rigorous conclude that homeopathic treatments fail to generate clinical effects that are different from those of placebo[164]’;
>
> ‘judging only from the restricted number of studies that really count (those conducted on a large clinical material and methodologically faultless) it is legitimate to conclude that the clinical effects of homeopathy are placebo effects[165]’;
>
> ‘after excluding methodologically inadequate trials and accounting for publication bias, homoeopathy produced no statistically significant benefit over placebo[166]’

Similarly the STC report states that:

> ‘The review which we consider the most comprehensive to date is that by Shang et al. … In our view, the systematic reviews and meta-analyses conclusively demonstrate that homeopathic products perform no better than placebos[167]’

That the clinical research evidence is univocally against the efficacy of homeopathic treatments is the accepted view in the Canonical Criticism.

[163] (Edzard Ernst, 2011a) p. 1007
[164] (Edzard Ernst, 2007) p. 2
[165] (Pandolfi, 2010) p. 148
[166] (Goldacre, 2007b) p. 1672
[167] (House of Commons Science & Technology Committee, 2010) paras 69-70.

3.1.5 Contesting the results of meta-analyses

Proponents of homeopathy contest the evaluation of the evidence given in the Canonical Criticism. Such arguments are put forward in addition to the more fundamental objections to the Canonical Criticism's view about what counts as evidence. Arguments contesting the evaluation of the evidence may be independent of the more fundamental arguments but not necessarily so; since as noted above a view about why PCTs are inappropriate may also explain negative results from PCTs. Considered below are those arguments which are independent, and do not presuppose some more fundamental objection: that is proponents who argue that the Canonical Criticism is technically incorrect in its assessment of the evidence from meta-analyses of randomised trials.

Most notably, controversy has built up around the most prominent meta-analysis of homeopathy: The 2005 *Lancet* paper by Shang et al. The method of the Shang et al meta-analysis is distinctive because it involved a matched comparison of placebo controlled trials of homeopathy with placebo controlled trials of conventional medicine (they were matched by condition and type of outcome). This allowed Shang et al to conclude:

> 'The effects seen in placebo-controlled trials of homoeopathy are compatible with the placebo hypothesis. By contrast, with identical methods, we found that the benefits of conventional medicine are unlikely to be explained by unspecific effects[168]'

In the paper 110 pairs of trials were identified and analysed. However, while all 110 pairs were used, for example, to produce the funnel plots estimating publication bias, the calculation of combined treatment effect used a subset of 'larger and higher quality' trials. The conclusion that homeopathic treatments are equivalent to placebo was based on eight trials of homeopathy (odds ratio: 0.88 [95% CI: 0.65-1.19]), and six trials of conventional medicine (OR: 0.58 [95% CI: 0.39-0.85]).

Many proponents of homeopathy have criticised the technical details of this analysis[169]. Three criticisms are often repeated:

[168] Shang et al 2005 p. 731

[169] In the STC report specifically, see: (House of Commons Science & Technology Committee, 2010) Ev. 136; Ev. 149; Ev. 169; Ev. 175; Ev. 193

Firstly, a number of authors take issue with poor reporting by Shang et al, most notably for failing to make it clear which 14 trials (8 homeopathy, 6 conventional) were used in the final meta-analysis[170] (This was subsequently corrected however[171]) – but also taking issue simply with the fact that only a subset of trials were analysed[172].

Second the analysis has been criticised because it lumps together homeopathic treatments for a range of different, heterogeneous, conditions. It is argued that this poses two problems: straightforwardly, it is supposed that if the analysed trials of homeopathy include results that are both true-positives for some conditions and true-negatives for other conditions, then a 'net' conclusion maybe drawn that homeopathy is ineffective for all conditions, when in fact it isn't[173]. Also a more sophisticated twist on this argument comes from critics who claim that the placebo effect may be larger (or at least relevantly more variable) in trials of homeopathy[174]. Defenders of Shang et al's result have noted that this was why the meta-analysis included matched pairs of trials[175]; and the to and fro continues when proponents contest whether the matching was appropriate[176].

Thirdly Shang et al are criticised for not performing any sensitivity analyses, that is to say, they do not test whether their conclusion holds if the set of trials analysed is altered[177]. As Ludtke and Rutten[178], and Rutten and Stopler[179] show, the result appears robust under the assumptions Shang et al made[180], however they go on to argue that there are other plausible analyses of the high quality trials which show a slight but significant effect of homeopathy. Although it has been further argued, on the contrary, that this is only apparent post hoc, given its dependence on particular statistical techniques (random effects meta-analysis, instead of meta-

[170] (Fisher, 2006; Fisher et al., 2005; Linde & W. B. Jonas, 2005)(D. Peters, 2005)
[171] (Shang, Jüni, et al., 2005)
[172] (Frass et al., 2005)
[173] (Chatfield, 2008; Dantas, 2005; Linde & W. B. Jonas, 2005; Rutten & Stolper, 2008)
[174] (Walach et al., 2005)
[175] (Shang, Jüni, et al., 2005; Paul Wilson, 2009)
[176] For example: (Fisher, 2006)
[177] (Fisher, 2006; Lüdtke & Rutten, 2008; Rutten & Stolper, 2008) and see also: (Kiene et al., 2005)
[178] (Lüdtke & Rutten, 2008)
[179] (Rutten & Stolper, 2008, 2009)
[180] (Lüdtke & Rutten, 2008) pp. 4-5 See specifically Figure 3 (and also Table 2)

regression) and perhaps dubious exclusion criteria (removal of the four homeopathic trials for muscle soreness; which are all negative)[181].

It is also worth noting a second prominent meta-analysis by Linde et al in 1997[182], again published in the *Lancet.* Linde et al reported a positive effect for homeopathy overall, however the data has been re-analysed multiple times[183]. The resulting argument made by opponents of homeopathy is that the original result was an overestimation; significantly, this is the conclusion reached by a re-analysis performed by the lead author of the original paper[184]. The reason for noting this is that Linde et al's meta-analysis is also commonly – and so the opponents of homeopathy claim, disingenuously – cited by proponents as evidence in favour of the efficacy of homeopathy. Indeed the omission of the re-analysis literature in the British Homeopathic Association's written submission to the STC Evidence Check was a particular point of contention in the STC's report[185].

3.1.6 The implausibility of homeopathy

A second line of argument (that is, in distinction to the line of argument given by (1) & (2); namely, about EBM and the primacy of randomised trials) in the Canonical Criticism concerns the plausibility of the claim that homeopathic treatments work[186]. The 'Implausibility Argument' is put forward in order to show that homeopathic treatments cannot work. The argument is based on inferring, from the implausibility of there being a mechanism by which homeopathic treatments could work, the claim that they don't work. The Implausibility Argument therefore puts forward mechanistic evidence against the efficacy of homeopathic treatment; as opposed to clinical research evidence, described above.

In support of the claim that there could not be a mechanism by which homeopathy treatments are efficacious opponents note the purported incompatibility of the similarity, small doses and dynamisation principles with

[181] (Paul Wilson, 2009)
[182] (Linde et al., 1997)
[183] See in particular: (Edzard Ernst, 2002; Linde et al., 1999)
[184] (Linde et al., 1999)
[185](House of Commons Science & Technology Committee, 2010) paras 66-68 & esp. 71, see also: Ev. 51-3
[186] (Baum, 2006; Baum & Edzard Ernst, 2009; Edzard Ernst, 2007, 2011a, 2011b; Hoffer, 2003; Holt, Gilbey, Colquhoun, Baum, & Edzard Ernst, 2011; Pandolfi, 2010, 2011; Sehon & D. Stanley, 2010; D. Stanley & Sehon, 2011; Vickers, 2000)

modern biomedicine[187]. Particular critical emphasis is placed on the counter-intutive notion that a substance could get more potent as it is diluted[188]. Often very strong claims are made about the implausibility of homeopathy (indeed, impossibility may be more accurate), for example:

> 'Those who claim that homeopathy is effective have enormous unexplained mysteries, and answering those mysteries would appear to require massive revision of standard chemistry and physiology… the balance is heavily against homeopathy[189]'.

> 'We understand that it cannot work through any mechanism that is in accordance with the known laws of nature[190]'

> 'If homeopathy worked the whole of chemistry and physics would have to be overturned[191]'

And most strongly:

> 'We think that a belief in homeopathy exceeds the tolerance of an open mind. We should start from the premise that homeopathy cannot work[192]'.

It is important to also note that the Implausibility Argument must be more than an argument from ignorance, otherwise it falls fouls of the well-known distinction between 'evidence of absence' and 'absence of evidence'. Just this point is made by Robert Wilson (Chairman, British Association of Homeopathic Manufacturers) in STC's first oral evidence session[193]. Wilson cites a number of examples of treatments that are known to be efficacious, but for which there is

[187] (Sehon & D. Stanley, 2010)(D. Stanley & Sehon, 2011)
[188] (Baum & Edzard Ernst, 2009)(Sehon & D. Stanley, 2010)
[189] (Sehon & D. Stanley, 2010) p. 281
[190] (Edzard Ernst, 2011b)
[191] (House of Commons Science & Technology Committee, 2010) Ev. 92
[192] (Baum & Edzard Ernst, 2009)
[193] (House of Commons Science & Technology Committee, 2010) Ev. 18 (response to Q79)

limited understanding of the mechanistic model behind them[194]. The Implausibility Argument does not present the view that homeopathic treatments might potentially be efficacious treatments, which lack a fully understood mechanism; rather it is an argument for the stronger view that the mechanistic implausibility of homeopathic treatments' efficacy is taken as evidence that they are not efficacious (as the quotes above illustrate). The claim made in the Implausibility Argument is that there is good evidence that there cannot be a mechanism, and not merely that it is not understood.

Within the Canonical Criticism and the STC Report the relationship between mechanistic evidence and clinical research evidence is not entirely clear. The Canonical Criticism places much more weight on the implausibility of the claim that homeopathic treatments work, compared to the STC Report. One might even argue that in the face of ambiguous clinical research evidence (that is to say, there are always likely to be methodological objections which enable the debate to be kept open) it is the Implausibility Argument which bears the most weight in the Canonical Criticism. This is not the view of the STC, however. Indeed there is a tension between this Implausibility Argument and the weight that the STC give to mechanistic evidence. According to the STC claims about the efficacy of treatments, inferred from knowledge of mechanisms, are supposed to possess minimal evidential weight:

> 'while we comment on explanations for how homeopathy works, it is not a key part of our Evidence Check…It is more important to know *whether* a treatment works—its efficacy—than *how* it works[195]'

> 'Lack of scientific plausibility is disappointing, but does not necessarily mean that a treatment does not work[196]'

The tension is between this idea that mechanistic evidence counts for little, and the strong claims that are made in the Canonical Criticism more widely about the

[194] More widely, general anaesthetics are perhaps the most frequently cited example, see: (Evers & Crowder, 2009)
[195] (House of Commons Science & Technology Committee, 2010) pp. 7 (para 18) original emphasis.
[196] (House of Commons Science & Technology Committee, 2010) p. 18 (para 65).

implausibility of a mechanism for how homeopathic treatments work. Interestingly, the STC also seem to endorse the same mechanistic claims made in the Canonical Criticism. They state:

> 'The principle of like-cures-like [what I have called the similarity principle: see §2.1.2] is theoretically weak. It fails to provide a credible physiological mode of action for homeopathic products. We note that this is the settled view of medical science[197]'

Additionally they state:

> 'We consider the notion that ultra-dilutions can maintain an imprint of substances previously dissolved in them to be scientifically implausible[198]'

The STC place almost no weight in this evidence: they infer very little from these statements. The STC report stands out because, unlike the wider critical literature, it does not deploy the Implausibility Argument. Ben Goldacre (journalist, phsycian) also expresses precisely this view in the first of the STC's oral evidence sessions, he states firstly that:

> 'the bottom line is it does not matter about the mechanism by which homeopathy is claimed to work or does not work; it does not[199]'

Then in answer to a follow-up question about how he knows this, he states the superiority of clinical research evidence:

> 'I think 200 trials which, taken collectively, showed that homeopathy pills worked no better than a placebo is very good evidence against homeopathy[200]'

[197] (House of Commons Science & Technology Committee, 2010) p. 16 (para 54).
[198] (House of Commons Science & Technology Committee, 2010) p. 17 (para 61).
[199] (House of Commons Science & Technology Committee, 2010) Ev. 20 (response to Q98)

The Implausibility Argument is clearly part of the Canonical Criticism of homeopathy: it is deployed widely in the literature. Interestingly however, it plays only a minor part in the arguments in the STC report.

3.1.7 Contesting the implausibility of homeopathy

The most direct challenge to the Implausibility Argument simply involves the claim that there is in fact a plausible mechanism by which homeopathic treatments could work; or at least that reasonable doubt can be cast on the Canonical Criticism's claim that there *cannot* be a mechanism. Sehon and Stanley have noted two further strategies[201]: first is a challenge to the idea that it is possible to give an explanation of the mechanism by which homeopathic treatments are efficacious; this relies on the claim that biomedical theory is incapable of capturing the way that homeopathic treatments work. Second the primacy of clinical research evidence is asserted; the argument here being (much like the STC's position) that mechanistic plausibility is not evidentially significant. Below I do not consider these two further challenges (the first amounts, in so far as it makes sense at all, to some notion that homeopathic treatments are 'complex', the second simply reiterates the tension already noted about the role that mechanistic evidence is supposed to play in the interpretation of EBM in the Canonical Criticism.).

Consider the argument that there are, in fact, plausible mechanisms for how homeopathic treatments could work. A number of authors have attempted to offer explanations of what are called 'memory of water' effects[202,203]. The Canonical Criticism focuses on the small dose principle as one of the key plausibility hurdles, however proponents of homeopathy argue that the relevant issue is not the contents of the dilutions, but their 'structure'[204]. It is for this reason that Peter Fisher, in the

---

[200] (House of Commons Science & Technology Committee, 2010) Ev. 20 (response to Q99)
[201] (Sehon & D. Stanley, 2010)
[202] See for example, the special issue of *Homeopathy* 2007:96(3) on the subject of the 'memory of water', notably (Chaplin, 2007)(Fisher, 2007) See also: (Fisher, 2010; Gold et al., 2008; Milgrom, 2006a)
[203] The STC notes the following concepts were all drawn upon in written submissions to their Evidence Check: 'electromagnetic properties, frequency imprinting, quantum physics and supra-molecular behaviour of water' (House of Commons Science & Technology Committee, 2010) para. 58
[204] (Fisher, 2010) see also: (Gold et al., 2008) p. 29

second oral evidence session of the STC evidence check, emphasised the dynamisation principle, since that principle supplies the link to structural changes in homeopathic dilutions; he states:

> 'You are inducing structural effects which may involve silica and which may involve dissolved oxygen molecules—it is not quite certain—but you can show that this water is different from water that is just shaken without the stuff [the substance being diluted] being in it[205]'

To support the emphasis placed on the structural properties of water (and therefore the dynamisation of homeopathic dilutions) proponents of homeopathy draw on evidence from materials science. Rustram Roy, making this same point about the importance of the structure of homeopathic dilutions, states that: 'The first law of material science is: ''properties are controlled mainly by structure, not by composition"'[206]'.

Consequently it is common for proponents of homeopathy to cite a literature of basic science research demonstrating these structural properties of water[207]. The result is that the Canonical Criticism's claim that homeopathic treatments could be efficacious only on pain of revising established knowledge about physics and chemistry is combated by counter-claims to the effect that:

> '[homeopathy presents a] challenge [to] the assumptions of high school chemistry, but not those of modern materials science[208]'.

Furthermore proponents of homeopathy point to uncontroversial claims within mainstream materials science that the physical chemistry of water is not well understood; as seen, for example, in the following quotation from an editorial by Philip Ball in *Nature*:

[205] (House of Commons Science & Technology Committee, 2010) Ev. 49 (response to Q156)
[206] (Gold et al., 2008) p. 29
[207] (Milgrom, 2006a, 2007)(Gold et al., 2008)(Chaplin, 2007)
[208] (Bell, 2005) p. 766

'no one really understands water. It's embarrassing to admit it, but the stuff that covers two thirds of our planet is still a mystery[209]'.

Again, the counter to this by opponents of homeopathy is that a significant plausibility hurdle remains to get from basic research investigating the physical chemistry of water to clinical effects in human beings[210]. However the most minimal claim made by proponents of homeopathy is consistent with this; since they may only aim to cast doubt on the Canonical Criticism's Implausibility Argument. That is to say, to shift the situation from 'evidence of absence' merely to absence of (or at least, contestable) evidence.

3.2 The policy debate

3.2.1 The Canonical Criticism

There is a similarly clear set of arguments that opponents of homeopathy make about the policy implications of their claim that homeopathic treatments are equivalent to placebo. The Canonical Criticism, with respect to the policy debate, is constituted by two ethical arguments:

(4) No Placebos argument: The provision of placebo treatments (and therefore homeopathy) necessarily involves deceiving, or violating the autonomy of, patients; as well as contributing to the medicalisation of the patients' complaints.

(5) Indirect Harm argument: The provision and state endorsement of placebo treatments (and therefore homeopathy) causes 'indirect harm' in so far as it creates the perception that they are efficacious medicines: because this perception may delay the treatment of serious conditions, or undermine public health advice.

Note first that while these two arguments are stable elements in the Canonical Criticism, their policy implications receive a more mixed treatment in the

[209] (Ball, 2008)
[210] (Sehon & D. Stanley, 2010; D. Stanley & Sehon, 2011)

literature. It is not immediately obvious whether these arguments support ending the provision of homeopathic treatment on the NHS, as the STC recommend[211], or merely altering aspects of its regulation as others have argued[212]; nor is it obvious how this support changes between the contexts of private-practicing homeopaths and homeopathy on the NHS[213].

Note second that the policy debate is not contested in the same way the evidential debate is. The arguments in the policy debate are premised on the notion that homeopathic treatments are equivalent to placebo[214]. For proponents of homeopathy the assumption that homeopathic treatments are placebo treatments is the foremost unsound premise in the debate. The arguments made by proponents often therefore reduce to issues already discussed in the evidential debate. They contend that homeopathy *does* work; therefore questions raised about the ethical and policy implications of placebo treatments are unconnected to a proper discussion of homeopathy, for example:

> '[opponents of homeopathy discuss] whether it is ethical for homeopaths to use a placebo if they know it is only a placebo. This debate is irrelevant; homeopaths know they are providing more than a placebo, both from their own clinical experience… and from the results of high quality studies[215]'.

In response to the STC's recommendation that the Government should re-examine the policy concerning the availability of placebo treatments on the NHS, the British Homeopathic Association similarly contends that:

> 'The committee has taken the rigid and incorrect view that homeopathy has been proven to be the same as placebo[216]'

and that

[211] (House of Commons Science & Technology Committee, 2010) paras 110-111
[212] (Hay, 2008)(Department Of Health, 2010) paras 42-43 (House of Commons Science & Technology Committee, 2010) Ev. 1
[213] (Edzard Ernst, M. H. Cohen, & J. Stone, 2004)
[214] (K. Smith, 2011) stands out for making the point most explicitly.
[215] (Ross, 2010) p. 297
[216] (British Homeopathic Association, 2010) p. 3

> '[its] recommendation hinges on the repeated assertion that homeopathy is a placebo… this view is not supported by scientific evidence[217]'

Considered below are replies made by proponents of homeopathy which contest other, non-evidential, points in the Canonical Criticism. Firstly however, the two arguments deployed in the Canonical Criticism, from which policy conclusions are supposed to follow, can be described as follows:

3.2.2 The 'No Placebos' argument

The No Placebos argument is an argument for the view that it is unethical to provide homeopathic treatments. Since first (according to the Canonical Criticism), homeopathic treatments are known to be placebos and second, it is unethical to provide placebos, then it is unethical to provide homeopathic treatments.

The important premise is that providing placebos is unethical. This is justified in a number of different ways. The STC closely follow the view put forward by Edzard Ernst in their second oral evidence session and give three reasons[218]: first, that placebo effects are unreliable; second, that placebo effects accompany all medical treatments, thus 'pure placebos[219]' are unnecessary; and third, that clinicians who provide placebos necessarily deceive patients.

When the STC elaborate on this however it is the third reason, and the notion of deception, that plays the central role[220], as is also the case in the wider literature[221]. The claim is that placebo treatments can only be effective by some deception of patients, by the clincian. Or alternatively, put in the form of a dilemma: that either, clinicians knowingly prescribe placebos and so are being deceptive; or

---

[217] (British Homeopathic Association, 2010) pp. 10-11

[218] See (House of Commons Science & Technology Committee, 2010) paras 36-37 & Ev. 46 (response to Q126). See also: Ev. 9, Ev. 102, Ev. 116

[219] A 'pure placebo' is a 'placebo' which is supposed to contain no active ingredients. As opposed for example to 'impure placebos' which may contain ingredients the are deliberately designed to produce, say, some side-effect; or alternatively contains some active ingredient whose effect is augmented by the placebo effect (A. K. Shapiro & Morris, 1978) p. 371

[220] (House of Commons Science & Technology Committee, 2010) paras 38 & 93-97

[221] Specifically in relation to homeopathy, see: (House of Commons Science & Technology Committee, 2010) para 38, see also paras 94-101, and (Baum & Edzard Ernst, 2009; Colquhoun, 2009a; Kavalier, 2011; K. Smith, 2011)

(and equally unethically), clinicians unknowingly prescribe placebos in ignorance of the available evidence, and so are being irresponsible[222]. Deception of patients by clinicians is held to be problematic because it is itself unethical (for traditional bioethical reasons, such as violating patients' autonomy etc) and moreover, it undermines the legitimacy of medical professionals. The STC report states that:

> 'deception arguably abuses the doctor-patient relationship [because]… when doctors prescribe placebos, they risk damaging the trust that exists between them and their patients[223]'.

The STC report is notable also because it constructs a similar argument from deception, but couched in terms of patient choice[224]. Here the claim is that it is an ethical requirement that patients should be able to make an informed choice about their treatments. Placebo treatments, because their effectiveness is necessarily connected to deception, must either be incompatible with patient choice, or if they are compatible, have their effectiveness reduced[225].

As noted above, it is less common for proponents of homeopathy to contest the specifically ethical aspects of the arguments presented by the Canonical Criticism; simply because they deny the crucial evidential premise, that homeopathic treatments are placebos. As might be expected, there is little direct engagement with the 'No Placebos' argument, for precisely this reason.

I now describe the second ethical argument found in the Canonical Criticism: the Indirect Harm argument.

3.2.3 The 'Indirect Harm' argument

The Indirect Harm argument is more specifically concerned with homeopathy, rather than with placebo treatments in general. Though there are different kinds of harm that the argument draws on, the claim is not that homeopathic treatments are themselves harmful or dangerous; instead the claim is that homeopathic treatments are indirectly harmful in so far as they put patients at

[222] (Goldacre, 2007b)
[223] (House of Commons Science & Technology Committee, 2010) paras 38 & 97
[224] (House of Commons Science & Technology Committee, 2010) paras 98-101
[225] (House of Commons Science & Technology Committee, 2010) para 101

risk in other ways[226]. Goldacre makes this point clearly in the first of the STC's oral evidence sessions:

> 'there are a number of harms that come [from homeopathic treatment], but none of them, you are absolutely correct to say, are direct physical harms. I do not believe that sugar pills are physically harmful[227]'

Similarly Stephen Evans, writing in the *British Medical Journal* sets out the structure of the argument:

> 'While this product [homeopathic treatments] may have no benefit, it probably has no direct harm either. But it may have major indirect harms—not only in individual patients who may not benefit from other effective remedies but also in a general sense by undermining the rational basis for medicine[228]'

In general there are two ways in which homeopathic treatment is supposed to cause indirect harm: firstly and most prominently, it is claimed that patients may be harmed as a result of the endorsement which homeopathic treatments receive from various institutions. The quote from Evans above picks out only one element of this general problem, namely patients foregoing effective treatment. See below for a second element (undermining public health advice). A second, more minor claim is also made for there being wider sociocultural harms that stem from the availability of homeopathic treatment.

Firstly: Harm stemming from the endorsement that homeopathic treatments receive is supposed to be caused by the perception that homeopathic treatments are efficacious. It is claimed that homeopathic treatments are tacitly endorsed when they are sold on pharmacy shelves[229], available through the NHS[230], protected by EU

[226] (Edzard Ernst, 2011b; Harris, 2011; Kavalier, 2011; Sample, 2008)(House of Commons Science & Technology Committee, 2010) paras 105-7 & 109, Ev. 9, Ev. 10, Ev. 12 (response to Q24 & Q25)
[227] (House of Commons Science & Technology Committee, 2010) Ev. 12 (response to Q25)
[228] (S. J. Evans, 2009)
[229] (House of Commons Science & Technology Committee, 2010) Ev. 9

legislation and regulated by the MHRA[231]. The Canonical Criticism makes the sociological claim that all of these tacit endorsements contribute to the perceived legitimacy of homeopathic treatments. For example, the STC report argues that the fact that homeopathic treatments fall under the remit of the MHRA exaggerates their credibility as medicines (in contrast, say, to the categories for cosmetics[232]). Similarly the STC argue that the NHS's constitution contains a commitment to providing effective treatments; thereby creating the reasonable expectation that homeopathy must be effective; they note that:

> 'When the NHS funds homeopathy it endorses it… [Since] the funding of drugs and treatments are made "following a proper consideration of the evidence", patients may reasonably form the view that homeopathy is an evidence-based treatment[233]'.

It is from this that the supposed harms are claimed to result. Firstly, the credibility homeopathic treatments have acquired, it is argued, may lead to delays in the diagnosis and treatment of serious conditions that require conventional medicine[234]. For example, the STC note:

> 'there is a risk that a patient whose symptoms improve following homeopathic treatment… may delay seeking proper medical diagnosis… for a serious underlying condition… [and] Patients who do not seek medical advice from properly qualified doctors run the risk of missing serious underlying conditions while they have their symptoms treated with placebo[235]'.

---

[230] (Harris, 2011; K. Smith, 2011) (House of Commons Science & Technology Committee, 2010) Ev. 10, Ev. 92, Ev. 102, Ev. 116, Ev. 131-2, see also Ev. 130-1

[231] (Deborah Cohen, 2009; Colquhoun, 2009b; S. J. Evans, 2009; Garattini & Bertelé, 2009; K. Smith, 2011) (House of Commons Science & Technology Committee, 2010) Ev. 9-10, Ev. 91-2, Ev. 102, Ev. 115, Ev. 118

[232] (House of Commons Science & Technology Committee, 2010) Ev. 184

[233] (House of Commons Science & Technology Committee, 2010) para 109

[234] (Hay, 2008)(S. J. Evans, 2009; Garattini & Bertelé, 2009; Harris, 2011; Sample, 2008; K. Smith, 2011) (House of Commons Science & Technology Committee, 2010) Ev. 27, Ev. 91, Ev. 101, Ev. 73

[235] (House of Commons Science & Technology Committee, 2010) paras 105-7

Also when homeopathic advice undermines public health advice, this may lead patients to eschew (rather than complement) conventional medicine[236]. There are a number of high profile cases, mentioned in Chapter 2[237], which the Canonical Criticism often emphasises[238], where there was a high risk of serious harm.

A second, less prominent, way in which homeopathy causes indirect harm is through wider sociocultural harms that result from the availability of homeopathic treatment[239]; medicalization being one such often mentioned harm[240]. Goldacre puts forward this argument most consistently[241]; he claims that:

> 'the very act of prescribing a pill carries its own risks: medicalisation, reinforcement of counterproductive illness behaviours, and promotion of the idea that a pill is an appropriate response to a social problem, or a modest viral illness[242]'.

Additionally, Tracy Brown (Managing Director, Sense about Science) put forward the view, in the first of the STC's oral evidence sessions, that the availability of homeopathy undermines the credibility of evidence-based medicine, she claims:

> 'There is also a broader harm to the public… We just lose, as a society, the dividing line, the ability to talk to people about the evidence behind their medicines, and I think that is a serious public health issue[243]'

[236] (Edzard Ernst, 2010b)(S. J. Evans, 2009; Goldacre, 2007b; Harris, 2011; Kavalier, 2011; K. Smith, 2011) (House of Commons Science & Technology Committee, 2010) Ev. 91, Ev. 101, Ev. 103

[237] Notably the undercover investigations of homeopathic treatments for malaria (Jones, 2006; Jones & Ghosh, 2011), and a controversial conference about the homeopathic treatment of HIV (Baum & Edzard Ernst, 2009; N. Cohen, 2007; Goldacre, 2007a; Mashta, 2009)

[238] (House of Commons Science & Technology Committee, 2010) Ev. 3, Ev. 7, Ev. 9

[239] (S. J. Evans, 2009; Goldacre, 2007a, 2007b; Sample, 2009a)(House of Commons Science & Technology Committee, 2010) Ev. 9, Ev. 12, Ev. 72

[240] See for example: (Conrad, 1992; Moynihan & Richard Smith, 2002)

[241] (Goldacre, 2007a, 2007b) (House of Commons Science & Technology Committee, 2010) Ev. 9

[242] (Goldacre, 2007b) p.1672

[243] (House of Commons Science & Technology Committee, 2010) Ev. 12

It is important to note however that the first set of arguments about the harms resulting from the endorsement of homeopathy by the NHS and MHRA is given the most emphasis. The sociocultural harms of homeopathic treatment plays a more minor role in the debate.

Taken in conjunction with the notion that homeopathic treatments are placebos, the conclusion that is drawn from these claims about the indirect harm homeopathic treatment causes has been summed succinctly up by Edzard Ernst:

> 'there is no good evidence to suggest that homeopathic remedies have any specific therapeutic effects and there is some evidence to show that homeopathy can cause harm. Thus its risk-benefit profile is negative[244]'.

3.2.4 Contesting the ethical arguments and their policy implications

As noted above, it is less common for proponents of homeopathy to contest the No Placebos argument; simply because they deny the crucial evidential premise, that homeopathic treatments are placebos. Since the Indirect Harm argument does not rely so heavily on the claim that homeopathic treatments are placebos, it is this argument which proponents more often contest. Below I note briefly challenges put by Lionel Milgrom, then address in more detail the way the DH disputes the STC's policy recommendations.

The Indirect Harm argument is addressed explicitly by Milgrom, who makes two counter-arguments[245]. Firstly he contests the degree to which homeopathic treatments might, in general, be indirectly harmful, he states:

> '[the Canonical Criticism claims] that those taking homeopathic remedies might forgo 'life-saving' drugs. This is a false perception: many who come to homeopathy do so only after conventional treatments have failed[246]'

[244] (House of Commons Science & Technology Committee, 2010) Ev. 27
[245] (Milgrom, 2009a)
[246] (Milgrom, 2009a) p. 257

Second Milgrom argues that, even if it is the case that homeopathy is indirectly harmful, it is also that case that conventional medicine is *more* harmful, and directly so[247]. In particular he cites a report by the UK House of Commons Public Accounts Committee[248], which notes the NHS' poor safety record and under reporting of medicines errors[249]. The point being that homeopathic treatment is *comparatively* harmless; and so Milgrom calls in to question the argumentative work that homeopathy's purported indirect harms are supposed to be doing[250].

Turning now to the policy conclusions that are drawn from the ethical arguments presented above, it is not altogether clear what the policy implications should be. Although the STC report is clear about its view:

> 'homeopathy should not be funded on the NHS and the MHRA should stop licensing homeopathic products[251]'

As one might expect, organisations such as the Society of Homeopaths[252] and the British Homeopathic Association [253] disagree with the STC's policy recommendations. Interestingly however the Department of Health (DH) also disagree (as briefly noted in Chapter 2), despite their broad agreement with the arguments made in the STC report.

Consider the DH response. In relation to the institutional endorsement of homeopathy, the DH report does indeed acknowledge that NHS funding of homeopathic treatments constitutes an endorsement of it[254], however they state further that:

---

[247] (House of Commons Science & Technology Committee, 2010) Ev. 96; See also: Ev. 136, Ev. 141. Note further the argument is similar to that of: (Illich, 1976)

[248] *A safer place for patients: Learning to improve patient safety (*2005) <http://www.nao.org.uk/publications/0506/a_safer_place_for_patients.aspx> Accessed Oct 2011

[249] (Milgrom, 2009a) See also: (O'Dowd, 2006)

[250] Interestingly, SSRIs for mild or moderate depression are arguably in the same evidential position as homeopathic treatments (Kirsch et al., 2008), and there have recently been calls questioning the prescription of SSRIs, for example (Middleton & Joanna Moncrieff, 2011). Indeed Milgrom also makes this point, see: (House of Commons Science & Technology Committee, 2010) Ev. 96

[251] (House of Commons Science & Technology Committee, 2010) para 157

[252] See for example: <http://www.homeopathy-soh.org/whats-new/latest-news/regu.aspx> Accessed Jan 2011

[253] See: (British Homeopathic Association, 2010)

[254] (Department Of Health, 2010) para 30

> 'There naturally will be an assumption that if the NHS is offering homeopathic treatments then they will be efficacious, whereas the overriding reason for NHS provision is that homeopathy is available to provide patient choice[255]'

Rather than remove homeopathic treatment from the NHS however, the DH take the view that, in light of the STC report, the reason for its availability needs to be made clearer, as part of the requirement that patients should be fully informed[256]. Explicitly the DH report states:

> 'providing appropriate information for patients should ensure that they form their own views regarding homeopathy as an evidence based treatment[257]'

The same kind of response is also made by the DH in relation to the MHRA's licensing of homeopathic treatments. The DH argues that regulation of homeopathic treatments is the best way to protect patients[258]. The DH explains their position by posing a problem:

> 'if homeopathic treatments were not subject to any kind of regulatory control consumers would not have access to such information or assurances [that those medicines are safe, manufactured to a high quality, and for specific purposes]. Conversely if regulation was applied to homeopathic treatments as understood in the context of conventional pharmaceutical medicines, these products would have to be withdrawn[259]'.

[255] (Department Of Health, 2010) para 9
[256] (Department Of Health, 2010) para 10 & 30, see also: paras 17-18
[257] (Department Of Health, 2010) para 30
[258] also they note, more fundamentally, that 'it is not open to the UK to set aside its obligations in European law to provide regulatory arrangements for homeopathic treatments' (Department Of Health, 2010) para 39
[259] (Department Of Health, 2010) para 37

The conclusion is therefore that the current situation (see §2.2.2), whereby homeopathic treatments fall into a (number of) special regulatory scheme(s), remains preferable to not regulating them as medicines. Indeed the DH state that:

> 'the fact that homeopathic medicinal products come within a regulatory scheme strengthens the ability of the MHRA to take regulatory action[260,261]

As well as arguing that regulation is the best way to protect patients, the DH also challenge other aspects of the Indirect Harm argument. Firstly they argue that the harms that are supposed to arise from the provision and regulation of homeopathic treatments are only significant in cases where serious conditions go untreated. The DH report downplays the significance of the supposed indirect harm argument, in much the same manner as Milgrom above. They state:

> 'we do not believe that this risk [of seeming to endorse homeopathy by providing and regulating it] amounts to a risk to patient choice or safety, nor do we believe that the risk is significant enough for the Department of Health to take the unusual step of removing PCT's [Primary Care Trust's] flexibility to make their own decisions[262]'.

[260] (Department Of Health, 2010) para 41

[261] The DH explain how EU legislation dealing with access to healthcare addresses this problem, quoting Recital 9 to directive 92/73/EEC (explained in Chapter 2), the DH report states: 'patients should be allowed access to the medical products of their choice, provided all precautions are taken to ensure the quality and safety of the said products' (Department Of Health, 2010) para 38.
They go on to note that Directive 2001/83/EC (see above also) establishes a registration procedure for homeopathic products which is simplified (as compared to conventional products) by exempting homeopathic treatments from the normal requirement to provide evidence of efficacy.

[262] (Department Of Health, 2010) para 47

Their point being that, even if homeopathic treatments are perceived as being endorsed by the MHRA (which the DH deny[263]), that does not have significant consequences. Consider for example the following statement:

> 'the main public health risk that can arise from homeopathic medicinal products is their misappropriate use in serious conditions[264]'

This risk is judged to be adequately managed because the medical claims that homeopathic treatments can make is, in virtue of the regulation that applies to them, limited to 'minor self-limiting conditions[265]'. To reiterate, the STC's 'Indirect Harm' argument began from the idea that patients may come to (falsely) believe that homeopathic treatments are legitimate and efficacious, because the MHRA regulation of them 'endorses' them. This endorsement was taken to be indirectly harmful because misleadingly classifying them as medicines can result in cases where more appropriate kinds of medical care should be, but in fact is not, sought.

In Contrast the DH position is that the harm that may arise from the perceived endorsement of homeopathic treatments should not be mitigated by ending their classification as medicines, but rather by making it clearer what the bounds of their appropriate use is. Consequently the DH report does note that:

> 'the MHRA is currently reviewing its guidance on the regulation of homeopathic treatments under the NRS [the National Rules Scheme] to ensure that the position on efficacy is clear[266]'

and notes further that:

> 'the MHRA will review the labelling requirements under the NRS [National Rules Scheme] to ensure that these deliver clarity as to the status of the products and their composition[267]'.

[263] Against the charge that MHRA regulation of homeopathic products suggests that they are efficacious medicines, the DH states: 'MHRA registration of products under appropriate regulatory schemes does not imply that the regulator is endorsing homeopathic products[263]'
[264] (Department Of Health, 2010) para 40
[265] (Department Of Health, 2010) para 40
[266] (Department Of Health, 2010) para 43

In the opinion of the DH then, the indirect harms of homeopathy are held to be too indirect to justify changes in current health policy.

3.3 Summary

The structure of the argument used in the homeopathy controversy has been described. The controversy can be divided into two distinct debates: firstly, about whether homeopathic treatment 'works', or not; and secondly about the policy implications of the first debate; should it be available, or not. In both cases, those who criticise homeopathy present a standard set of arguments: what I have called, the Canonical Criticism of homeopathy.

With regard to the evidential debate the view presented in the Canonical Criticism is that the standards of 'evidence-based medicine' provide the framework for assessing whether homeopathy works. The key question to ask is whether homeopathic treatment is efficacious, rather than merely effective. The way to answer this question is by determining whether homeopathic treatments outperform placebo in randomised trials. Moreover, the Canonical Criticism holds that the best available evidence from such trials shows that homeopathic treatment is equivalent to placebo. Additionally, a second way to answer this question is by evaluating whether it is mechanistically plausible that homeopathic treatments could be efficacious. The Canonical Criticism holds that it is too implausible to expect homeopathic treatment to be efficacious, however the STC do not place significant weight in the Implausibility Argument. The relationship between the Implausibility Argument and the view that efficacy is best determined by placebo controlled trials is unclear.

In various ways proponents of homeopathy contest both the evaluation of the evidence and the account of what counts as evidence, presented in the Canonical Criticism. For instance it is argued that homeopathic treatment should be thought of as a 'complex intervention', which challenges the appropriateness of placebo controlled trials for the determining whether it works, furthermore it is argued that existing trials of homeopathic treatment do not, as the Canonical Criticism presents them, paint a negative picture of its efficacy. In addition it is also argued that the

[267] (Department Of Health, 2010) para 44

efficacy of homeopathic treatment is not mechanistically *impossible*; and therefore should be the subject of further open-minded debate. More generally proponents claim that EBM, as it is embodied in the Canonical Criticism (and perhaps also as it is perceived in medicine more widely) is synonymous with the reification of randomised trials. Such a view is accused of being based on unsophisticated evidence hierarchies that hold up the randomised trial as the one and only 'gold' standard of evidence. This, proponents of homeopathy claim, ignores important contributions from evidence ranked lower down the hierarchy. More fundamentally, proponents of homeopathy question the epistemological coherence of hierarchies at all. The challenge brings into focus the two questions of both how EBM is interpreted (proponents of homeopathy claim, as a matter of fact, the interpretation is naïve) and how it should be interpreted (they claim that the interpretation should be more sophisticated).

With regard to the policy debate there are again a canonical set of arguments put forward by opponents of homeopathy. Significantly the Canonical Criticism puts forward two ethical arguments. The first argument, the No Placebos argument, is that the provision of homeopathic treatment necessarily involves deceiving, or violating the autonomy of, patients; and is therefore unethical. The second argument, the Indirect Harm argument is that homeopathic treatment is possibly harmful (albeit indirectly) but offers no benefit; and is therefore unethical. From these two arguments, the conclusion drawn in the Canonical Criticism is that homeopathic treatment should not be available to patients, and should not be treated as if it were a legitimate medical treatment. The idea is that *outperforming placebo* is the measure of whether a treatment works, provides the fundamental premise for the ethical aspect of the controversy about homeopathy. The identity of homeopathic treatments with placebo medicines de-legitimises those medicines on account of the potential, uncompensated, harm they may cause and the dubious ethics associated with their very availability.

Again, these arguments are challenged by proponents of homeopathy. Mostly by denying the crucial evidential premise that homeopathic treatments are placebo treatments, but more significantly by questioning whether homeopathic treatment is harmful to any significant extent, and also by comparing those supposed harms to the direct harms that conventional medicines cause. Indeed it is notable that the Department of Health take this view; and therefore hold that the provision

of homeopathic treatment is ethically acceptable, if patients are properly informed about when, and when not, homeopathic treatment is appropriate.

CHAPTER 4

4. Summary of Part One and the questions to be addressed in this thesis

There are more arguments in the controversy than could be examined in detail in this thesis. Part One aimed to describe the controversy's breadth over both evidential and policy debates. The focus of this thesis will be the two key themes which run through and structure those debates: namely, evidence-based medicine and placebos.

The question of interest here is how strong the dual conceptual foundations of EBM and placebo really are in the Canonical Criticism. *EBM set the epistemological framework for evaluating treatments: placebos set the epistemological and ethical standard that legitimate treatments must pass.* Part Two examines EBM in more detail: Part Three, placebos. This thesis asks whether these two concepts 'do the work' which is expected of them in the Canonical Criticism.

More specifically, this thesis will address the follow questions:

(1) How does the interpretation of EBM used in the Canonical Criticism compare to the way EBM is interpreted in the medical literature?
(2) How should EBM be interpreted; specifically, what role should the mechanistic evidence put forward in the Implausibility Argument play in debates about homeopathy?
(3) How should the alleged complexity of homeopathic treatment affect the view that one can measure their efficacy in placebo controlled trials?
(4) What is the ethical significance of placebo comparisons? Why is it that outperforming placebo should be thought to affect the permissibility of providing a treatment?

The Canonical Criticism is supposed to supply one with reasons for holding particular beliefs about the efficacy of homeopathic treatments and the role those medicines should play in a rational healthcare system. This thesis takes an interest in the structure of those reasons. Firstly because of what might be gained in terms of making a stronger case, or understanding better the cases being made, for or against

the availability of homeopathic treatment. Secondly because of what can be gained in terms of our understanding of the nature of medical evidence in general.

4.1 Introduction to Part Two

Proponents of homeopathy often single out the concept of 'evidence-based medicine' for criticism, because it is a concept on which the Canonical Criticism draws heavily. Problems arise when one considers the particular interpretation of EBM that is offered in the Canonical Criticism. Proponents of homeopathy are keen to point out the naivety of the interpretation of EBM that the Canonical Criticism draws on. The Canonical Criticism is accused of reifying evidence from randomised trials (See Chapter 3). This introduces a tension in the way that opponents of homeopathy construct the evidential debate. The STC report, for example, seemed committed to an interpretation of EBM that holds that mechanistic evidence possesses little evidential weight; but it also asserted that homeopathic treatments cannot work because they have a grossly implausible mechanism. It is these issues around the proper interpretation of EBM that Part Two will examine in more detail.

The first question to consider is what interpretation(s) of EBM are offered in the medical literature? The purpose of asking this question is to evaluate the extent to which EBM provides an adequate foundation for the arguments put forward in the evidential debate, in the Canonical Criticism.

To anticipate: It will be argued in Part Two that this evaluation is somewhat less straightforward than one might expect. The key claim in Part Two will be that the EBM literature is unhelpfully unclear about the interpretation that *is* held. I will argue that examination the medical literature presents a set of basic arguments for EBM, from which only very weak conclusions are drawn. This has led to criticism of EBM and to ‘evidence hierarchies’ seeming to do much of the epistemological work in the EBM philosophy of evidence. It has given rise to (accusations of, at the very least) an interpretation that ranks different kinds of evidence as *categorically* better or worse than others. Importantly, this Categorical Interpretation is the interpretation of EBM that is offered by the STC in their report, and which proponents of homeopathy attack.

Part Two will also briefly examine the question of what interpretation *should* be held. On one very plausible account, offered recently by philosophers of science,

EBM can be interpreted in a way that promises to resolve the tension between the STC's interpretation of EBM and the mechanistic argument that is put forward in the Canonical Criticism for the implausibility of homeopathy. This analysis provides the tools for a re-evaluation of the arguments put forward by the STC and the Canonical Criticism; which will be the subject of Part Four.

PART TWO: EVIDENCE BASED MEDICINE

CHAPTER 5

5. What is Evidence-based Medicine?

The purpose of this chapter is to introduce EBM, and the arguments that have been put forward in favour of the EBM philosophy of evidence. That is, the arguments for why one should trust clinical research evidence over clinicians' experience or mechanistic evidence. This chapter also briefly describes some of the criticisms that EBM has faced, and the role that 'evidence hierarchies' play in the EBM philosophy.

5.1 Origins and definition

In order to give an account of what EBM is supposed to involve and the aspects of EBM that matter for this discussion it is helpful to consider EBM's intellectual origins, in particular its emergence from the discipline of clinical epidemiology[268]:

From among a range of authors making similar points in the 1960s, two important series of articles published in the Annals of Internal Medicine by the US clinician and epidemiologist Alvan Feinstein can be picked out[269]: A four part series 'The Scientific Methodology in Clinical Medicine[270]' in 1964 and the three part series 'Clinical Epidemiology[271]' in 1968[272]. The importance of these series is the approach to clinical practice that they advocate. In the first, Feinstein argues that clinical data can be valuable, just as laboratory data is, if it is collected systematically and rigorously. He claims that the heterogeneity of clinical practice demands, rather than prohibits, a scientific approach on a par with laboratory research. In the second, he characterises 'clinical epidemiology' as the application of epidemiological methods to the study of *clinically* defined populations; so as to be able to generate results that will improve clinical practice.

[268] On this topic see especially: (Daly, 2005)
[269] See: (D. L. Sackett, 2002). (Daly, 2005) ch. 2 also picks out the contributions of Henrik Wulff and Kerr White.
[270] (Feinstein, 1964a, 1964b, 1964c, 1964d)
[271] (Feinstein, 1968a, 1968b, 1968c)
[272] These were followed by his books *Clinical Judgement* in 1967 and *Clinical Epidemiology* in 1985.

During the 1960's and '70's, influenced by and in parallel with Feinstein's work, David Sackett and colleagues developed these ideas at McMaster University in Canada[273]. Members of the Clinical Epidemiology & Biostatistics department at McMaster University published a number of important articles during this time, setting out the principles and application of clinical epidemiology[274]. Further building on this, in 1985 Sackett and colleagues published the text book 'Clinical epidemiology: A basic science for clinical medicine[275]'. In this book they stated the rationale of their approach in the following terms:

> 'All of us believed we were practising the Art (derived from the beliefs, judgements and intuitions we could not explain)…[But there is] *a Science to the Art of medicine*…[and applying] epidemiologic principles (plus a few more from biostatistics) to the beliefs, judgements and intuitions that comprise the art of medicine might substantially improve the accuracy and effectiveness of diagnosis and prognosis, the effectiveness of management, the efficiency of trying to keep up to date, and, of special importance, the ability to teach others how to do these things[276]'

The key insight here, like in the earlier work of Feinstein, is that elements of medical care that typically rely on evidence derived from the expertise of clinicians can be improved through being informed by more rigorous evidence. As the italicised section of the above quote makes explicit, clinical epidemiology is an attempt to augment the knowledge that clinicians rely on when making decisions about the treatment of individual patients.

---

[273] For example one key publication is: (D. L. Sackett, 1969)

[274] For example, the 'How to Read Clinical Journals' series published in the Canadian Medical Association Journal introduced clinicians to techniques for critically appraising published research, that is, techniques for evaluating and then translating research reported in the medical literature into clinically useful information. See:(Department of Clinical Epidemiology & Biostatistics McMaster University, 1981a, 1981b, 1981c, 1981d, 1981e). Note aside the tongue in cheek continuation of this series (Redelmeier & Shumak, 2003; Redelmeier, Shuchman, & Shumak, 1998; Shumak & Redelmeier, 2000, 2004)

[275] There are of course other examples of books of this sort, which consolidate the same arguments – another notable example is: (R. H. Fletcher, S. W. Fletcher, & E. H. Wagner, 1982)

[276] (D. L. Sackett, R Brian Haynes, & Peter Tugwell, 1985) p. ix [my emphasis]

From this kind of work EBM emerged as a named concept at the beginning of the 1990's[277,278,279]. The close link between evidence-based medicine and clinical epidemiology provides the first handle on what the EBM philosophy of evidence involves[280]: EBM, like clinical epidemiology, is primarily about the use of systematic research evidence to inform clinical practice. The frequently quoted statement from Sackett et al in the *British Medical Journal* further adds to this general idea:

> 'Evidence based medicine is the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients[281]'.

It is important to make a preliminary refinement of this general idea: there are at least two independent ideas that can be distinguished[282,283]. This is highlighted by asking what 'evidence' is supposed to be evidence *for*. Consider:

(1) EBM as an account of what constitutes evidence *for clinical decisions*

EBM in this sense is a method for solving particular practical problems. The focus is on the skills and methods clinicians should use to apply relevant evidence to a particular situation. EBM in this sense provides an account of how to find, assess and act on evidence. Typical 'definitions' of EBM that have this sense in mind are the following:

---

[277] Specifically, it first appeared in: (G. H. Guyatt, 1991)

[278] Jeanne Daly summarises the emergence of EBM as follows: 'The clinical epidemiologists polished their product more highly and marketed it under an attractive new label, evidence-based medicine. Evidence-based medicine was explicitly promoted in workshops and in an extensive literature as the best response to everyday problems encountered in clinical care' (Daly, 2005) p. 211

[279] By the mid-1990s EBM had become an established concept. In 2008, nearly twenty years after its first appearance in the medical literature, Montori and Guyatt were able to state with good reason that: 'the influence of EBM has been widely recognised in both lay publications (eg. The New York Times listed EBM as one of its ideas of the year in 2001) and in the academic press (eg. The BMJ listed EBM as one of the 15 greatest medical milestones since 1840)' (Montori & G. H. Guyatt, 2008) p. 1814

[280] This fact is presented as if it should be a surprise in, for example, (La Caze, 2009)

[281] (D. L. Sackett, Rosenberg, Gray, R Brian Haynes, & W. S. Richardson, 1996) p. 71

[282] See also: (Tonelli, 1998) p. 1235

[283] More generally, EBM is used in a variety of ways: it can be, for example, a 'strategy', an 'epistemological idea', and even a 'social movement'. See: (Kristiansen & Mooney, 2004) also: (Pope, 2003)

> 'Evidence based medicine is about asking questions, finding and appraising the relevant data, and harnessing that information for everyday clinical practice[284]'

> '[EBM is about] integrating individual clinical expertise with the best available external clinical evidence from systematic research[285]'

> 'EBM requires clinical expertise for producing and interpreting evidence, performing clinical skills, and integrating the best research evidence with patient values and circumstances[286]'

(2) EBM as an account of what constitutes evidence *for medical claims*

EBM, in this sense, is a view about what counts as evidence, the epistemic merits of different kinds of evidence, and the ways those different kinds of evidence relate to each other – EBM considered as an epistemological thesis. This is EBM considered less as a strategy for action, but as a method for determining truth. What I mean by 'medical claims', in distinction to the more straightforwardly understandable 'clinical decisions', are claims which are made, for example, about the efficacy of a drug for some condition, the accuracy of a diagnostic test, the likelihood of a serious side-effect. To illustrate, a typical 'definition' that focused on this sense would be:

> 'Evidence-based medicine de-emphasises intuition, unsystematic clinical experience, and pathophysiological rationale as sufficient grounds for clinical decision making and stresses the examination of evidence from clinical research[287]'

[284] (Rosenberg & Donald, 1995) p. 1122
[285] (D. L. Sackett et al., 1996) p. 71
[286] (Howick, 2011) p. 183
[287] (Evidence Based Medicine Working Group, 1992) p. 2420

> 'Evidence-based medicine emphasises the need to move beyond clinical experience and physiological principles to rigorous evaluations of the consequences of clinical actions[288]'

These two aspects of EBM are complementary. EBM in sense (1) is about determining the best course of action. That is, exercising clinical expertise to decide how to treat a patient. For example, deciding whether a patient with type 2 diabetes should be treated with either metformin or sulphonylureas (e.g. gliclazide); depending on, say, their weight and renal function. The question which EBM in sense (1) attempts to provide an answer to is: "what should I do to help my patient?". EBM in sense (2) is about determining the evidence that can be used as an ingredient in those decisions. EBM in sense (2) offers an answer to the question "what counts as good evidence for this claim?", or "is there good evidence for this claim?". (2) supplies the epistemological content that underwrites (1). It explains, for example, why and when some piece of evidence provides good evidence for a medical claim.

Notice that the process of 'critical appraisal[289]' cuts across this distinction. The techniques of critical appraisal concern both the assessment of the quality of evidence (is this good evidence?), and the applicability of evidence when making a decision (can I use it?)[290]. In what follows the primary concern is with EBM in sense (2). Since, as was shown in Part One, EBM is enrolled in the homeopathy controversy in this more epistemological sense; that is, as an authority on what counts as evidence, and what kinds of evidence are reliable. EBM is a named resource specifically in the more fundamental debates about what counts as evidence; about why efficacy matters; and why placebo-controlled trials are best placed to provide the most reliable answer to the question of whether homeopathy works.

In the next section I outline the basic arguments that have been put forward for the EBM philosophy of evidence.

---

[288] (Oxman, D. L. Sackett, & G. H. Guyatt, 1993) p. 2093

[289] The process is explained, for example, in: (Crombie, 2008; Trisha Greenhalgh, 2006; G. H. Guyatt & Drummond Rennie, 2002)

[290] This is clear in, for example: (G. H. Guyatt & Drummond, 2002a; D. L. Sackett, W. S. Richardson, Rosenberg, & R Brian Haynes, 1997; Straus, W. S. Richardson, Glasziou, & R Brian Haynes, 2005)

5.2 The basic arguments for evidence-based medicine

If the literature is viewed as an attempt to elucidate the precise details of the EBM philosophy of evidence, then I claim that those details are not clear. A general sketch of the EBM view and its supporting arguments however are both relatively clear in the literature and highly plausible[291]; and indeed, they have been since the term first appeared in 1991[292].

The quotation given above from the 1992 JAMA paper presents a typical 'definition' of EBM qua epistemological thesis. It provides an excellent starting point for an explanation of the general idea behind EBM; again:

> 'Evidence-based medicine de-emphasises intuition, unsystematic clinical experience, and pathophysiological rationale as sufficient grounds for clinical decision making and stresses the examination of evidence from clinical research[293]'.

This quotation is helpful for both historical and philosophical reasons: Its historical virtue is that it is one of the earliest uses of the term evidence-based medicine; since it comes from a 1992 paper by the, then relatively newly-formed, Evidence-based Medicine Working Group. Its philosophical virtue is that it is epistemologically explicit: it picks out three kinds of medical evidence (experience, mechanisms and research) and ranks them. It tells us that results from clinical research must be given more 'emphasis' than clinicians' experience or mechanistic theory.

The following sections explain why it is that the results from clinical research are supposed to be emphasised over experience and mechanisms. The argument itself is relatively clear[294]: the conclusions that are drawn however are surprisingly weak.

---

[291] At the most general level no one contests the idea that we should base medicine on evidence – and indeed, it has become relatively common to make just this observation at the beginning of books and papers about EBM.

[292] Gordon Guyatt introduced the term evidence-based medicine in 1990 (first used in the academic literature in (G. H. Guyatt, 1991)), after criticism of the term 'scientific medicine'. See: (Daly, 2005) p. 85

[293] (Evidence Based Medicine Working Group, 1992) p. 2420

[294] It shares much in common with the arguments that motivated the approach of clinical epidemiology. These ideas are clearly present in, for example: (S. Fletcher, E.

5.2.1 De-emphasising Experience

As Sackett et al noted in the quotation from *Clinical Epidemiology* above, a clinician's experience equips them with 'beliefs, judgements and intuitions we [they] could not explain'. 'Experience' here refers to those beliefs, judgement and intuitions that are acquired vicariously in day-to-day practice, from mentors, or from casual reading of the literature, etc. (as opposed to beliefs, judgements and intuitions acquired from the explicit examination of the available evidence).

An initial point to make is that, in so far as EBM 'de-emphasises' these beliefs, judgements and intuitions – in short: experience – the concern is with the evidential role they play. The important point is that it can be formulated propositionally: that is to say, the concern is with tacit experience which can be *evidence*[295]. This is not meant to imply that other roles they play should be de-emphasised too, though below it will be noted that one point on which EBM is criticised is that it does seem to have this implication[296].

The reason given for the de-emphasis of the evidential role of experience is that it is unsystematic. This is taken to be a problem because its unsystematic nature makes it too sensitive to bias and error[297]. The point often made is that clinical experience is idiosyncratic and heterogeneous: it consists of, as was noted above for example, judgements about which prior cases were similar, conversations with colleagues, perhaps a small sample of journal articles that can be recalled. In an attempt to explain the kind of evidence that clinical experience aims to provide,

---

Wagner, & R. Fletcher, 1996; D. Katz, 2001; D. L. Sackett, G. H. Guyatt, P Tugwell, & R Brian Haynes, 1991)

[295] I take evidence to be propositional – but note also that the relationship between 'knowing how' and 'knowing that' proposed by Gilbert Ryle has received extensive treatment in the philosophical literature: See especially (J. Stanley & T. Williamson, 2001), who argue for the view that knowing-how is a species of knowing-that. I expect that a more detailed examination of the relationship between the evidential and non-evidential aspects of clinical experience to be more complicated than the simple account given here.

[296] That experience also plays other, non-evidential roles, can be seen for example in the range of diagrams that have been produced in order to carve out a synthesising/expertise role for experience (See for example: (Lambert, 2006) figs 1 & 2). And note that once we start asking more precise questions about the roles that experience can usefully play in clinical practice, there is a need for finer grained distinctions between different elements that are here lumped together under the term 'experience'. See in particular (Howick, 2011) ch. 11

[297] The idea is expressed most clearly in: (Evidence Based Medicine Working Group, 1992; G. H. Guyatt et al., 2000)

David Katz has claimed that: '[clinicians] are all de facto clinical epidemiologists[298]'. But as Katz notes, clinical experience falls short precisely because it only draws on vicarious experience of an ill-defined population of unreliably-similar cases. The problem here, so the arguments goes, is that the lack of reliability which this idiosyncrasy implies makes experience evidentially weak. As the EBM Working Group state in their 1992 JAMA paper: 'In the absence of systematic observation one must be cautious in the interpretation of information derived from clinical experience and intuition, for it may at times be misleading[299]'. In fact, there is a range of evidence to back-up this kind of argument, which applies to fields of expertise generally (as well as to medicine). For instance there is much research from the psychological literature showing the extent to which experts' judgements are prone to cognitive bias[300].

It is important to note that no precise conclusion is drawn from this kind of argument. The statement above by the EBM Working Group is telling: they only draw the minimal and weak conclusion that experience may mislead, without detailing the circumstances when or to what extent.

5.2.2 De-emphasising Mechanisms

Mechanistic evidence clearly provides something quite different to evidence from clinical research, or from a clinicians' experience. Experience was presented above as a kind of botched clinical research: clinicians may be 'de facto clinical epidemiologists', but the problem is that they are bad ones. Mechanistic evidence is not like this: it is not poor quality clinical research. Instead of inferring some therapeutic effect from facts about a comparison between groups, the inference is made with reference to facts about the causal structure of some relevant system. Mechanisms in biology have been a subject of philosophical interest[301], however at this level of explanation it is sufficient to note two points. First, that 'mechanisms' can be construed broadly to refer to systems with parts that interact in regular ways. Second, that 'mechanistic evidence' – or as it is sometimes called

[298] (D. Katz, 2001) p. xii
[299] (Evidence Based Medicine Working Group, 1992) p. 2421
[300] To give one very recent example, see: (Berghmans & Schouten, 2011; Imam, 2011; S. J. Newell, 2011)
[301] See for example: (Glennan, 1996, 2002; Machamer, Darden, & Craver, 2000; Russo & J. Williamson, 2007)

'pathophysiological[302]' evidence or 'basic science[303]' evidence – is supposed to justify the inference from the presence of some intervention, via knowledge of the appropriate mechanism, to the presence of therapeutic effects.

The reason for de-emphasising mechanistic evidence is straightforward. As with clinical experience, the point is that mechanistic evidence is rarely reliable. The case against experience was made by reference to the possibility of error and the case against mechanistic reasoning is the same. A reliable mechanistic bridge between an intervention and therapeutic effects is difficult to establish; as Adam La Caze illustrates succinctly:

> 'Much is unknown in clinical science. Pharmacological/pathophsyiological mechanisms sometimes predict patient outcomes, and sometimes they don't; in any particular instance, it is often unknown which will be the case until applied clinical studies have been conducted[304]'

Unlike the case against experience however, the argument for de-emphasising mechanistic reasoning is often made by referring to a set of well-known cases, where such reasoning was demonstrably mistaken[305,306]. The classic examples are: Antiarrhythmic drugs given after myocardial infarction[307], Hormone replacement therapy in vascular prevention[308], Fluoride treatment of osteoporosis[309], and high-doses of aspirin for carotid endarterectomy[310]; were all thought, mechanistically, to be likely to be beneficial but in fact were shown to be harmful. In the other direction, the treatment of congestive heart failure with beta-blockers[311] for example, appeared

[302] The point is simply to include any of the non-clinical sciences which deal with mechanisms (construed in the broad sense above) – such as immunology, physiology, or pharmacology – and which might usefully inform clinical medicine.
[303] (La Caze, 2011)
[304] (La Caze, 2011) p. 88
[305] (La Caze, 2011) p. 85-6 makes this same observation that the argument for trusting randomised trial results over basic science is premised on: 'a number of prominent case studies'.
[306] Combinations of these examples are drawn upon in papers such as: (G. H. Guyatt et al., 2000) and (Rothwell, 2005) See also, (Howick, 2011) Ch. 10 Appendix
[307] (Echt et al., 1991; The Cardiac Arrhythmia Suppression Trial (CAST) Investigators, 1989)
[308] (Rossouw et al., 2002)
[309] (Meunier et al., 1998; Riggs et al., 1990)
[310] (Taylor et al., 1999)
[311] (Eichhorn & Bristow, 2001) (G. H. Guyatt et al., 2000)

mechanistically to be likely to be harmful but was later shown in randomised trials to be beneficial.

Again the conclusion that this argument supports is not precisely described. The EBM Working Group again illustrate the point: '[mechanisms] are necessary but insufficient guides for clinical practice... The rationales for diagnosis and treatment, which follow from basic pathophysiological principles, may in fact be incorrect[312]'.

The fact that mechanistic evidence 'may... be incorrect' is unhelpful: the issue is when and why mechanistic evidence can provide good or bad evidence. The EBM literature is not clear on these details[313]. Neither the nature of the relationship between mechanistic evidence and other kinds of evidence, nor the details of how clinical practice can be guided by mechanistic evidence are precisely specified. The implications of the argument for 'de-emphasising' mechanisms are not drawn out in detail, as with the argument for 'de-emphasising' experience. The conclusion drawn is weak.

5.2.3 Stressing Clinical Research

The arguments presented above are supposed to supply reasons for de-emphasising evidence which comes from experience and mechanisms; at the same time those arguments thereby show why evidence that comes from clinical research should be stressed. The advantage of evidence obtained from clinical research is that it lacks the disadvantages of experiential or mechanistic evidence. That is to say, clinical research is supposed to be less prone to be bias and error than either experiential or mechanistic evidence[314].

This is not controversial. Measures for bias minimisation in clinical research are well-known and straightforward[315]: for example, performing controlled comparisons perhaps involving 'placebos[316]', randomising the participants[317] and, along with the investigators, blinding them. The design of clinical studies is its own

[312] (Evidence Based Medicine Working Group, 1992) p. 2421
[313] (La Caze, 2011)
[314] I have picked out this 'bias minimisation' theme in order to explain the basic arguments for the EBM view. (Borgerson, 2009) picks out in her discussion of arguments for ranking different kinds of evidence in hierarchies, an almost equivalent 'causation-establishing' theme.
[315] For a thorough investigation see: (Howick, 2011)
[316] See Part Three. See also: (Howick, 2009, 2011)(Nunn, 2009a, 2009b)(Turner, 2012)
[317] The epistemic virtue of randomisation has been a topic of particular interest to philosophers in recent years, (La Caze, 2009; La Caze, Djulbegovic, & Senn, 2011; Worrall, 2007a, 2007b, 2010) see: also: (Papineau, 1994)

field and the task of determining which measures are most appropriate for a given research question is a design issue. The point to make here is that at least one of the purposes of any kind of empirical investigation is to test hypotheses and distinguish between theories. In contrast to experience and mechanistic evidence, clinical research involves setting up situations in order to do this systematically and reliably.

Proponents of EBM also claim the advantages of clinical research are demonstrated in a set of classic examples. For instance, the examples given above concerning mechanistic reasoning are repeated in this context, since as well as being examples where mechanistic reasoning was at fault they are also examples where the correct view was revealed in randomised trials.

Additionally there are other examples such as: foetal heart rate monitoring[318], High-dose oxygen treatment for neonates[319], extracranial-intercranial bypass surgery[320], all of which, on the basis of prior unsystematic evidence, were thought to be beneficial, but which randomised trials showed to be harmful.

Similar examples can be found where an intervention previously thought to be ineffective or harmful has been shown to be beneficial, as a result of randomised trials being performed. For example, the unexpected reduction in mortality gained from giving beta-blockers to patients with congestive heart failure[321] (noted above) and the reduction of infant respiratory distress (and resulting problems) gained by giving steroids to mothers at risk of premature labour[322,323].

One must be careful about the kind of work one expects these examples to do, however. Providing examples of cases where clinical research has found a treatment benefit as justification for the claim that they are better equipped to find such benefits begs the question. The problem is that disagreement between clinical research and other methods is not itself evidence for the superiority of clinical research. This question-begging has been noted previously[324]. What is needed therefore is some independent reason for thinking that clinical research is a better

[318] (Worrall, 2007b)
[319] (Silverman, 2004)
[320] (The EC/IC Bypass Study Group, 1985)
[321] (G. H. Guyatt et al., 2000)(Eichhorn & Bristow, 2001)
[322] (Rosenberg & Donald, 1995)
[323] This last example has significance for another reason. The reduction in respiratory distress and other complications was found as the result of a meta-analysis of a number of less conclusive randomised trials. The odds-ratio plot from this analysis, clearly showing an overall beneficial effect, was adopted as the logo for the Cochrane Collaboration. See: (Chalmers, D. L. Sackett, Silagy, & Maynard, 1997)(Cochrane Collaboration, 2009)
[324] (Grossman & Mackenzie, 2005)(La Caze, 2009)

measure of treatment effects. That independent reason comes from the bias-minimisation idea which, as has been noted, underlies the arguments described above. Clinical research is supposed to offer a better measure of treatment effects because it includes measures to reduce, and therefore is less prone to, the bias and error found in less systematic methods.

In summary I claim that the arguments here are not controversial; however the conclusions that are based on them are very weak. There is little exploration of the relationships between different kinds of evidence. Conclusions are drawn at a level of generality that leaves phrases such as 'de-emphasises' or 'stresses' to do much of the implicit work. One might be suspicious that, if the argument is taken seriously, it is an argument for the view that one should stress, *above all else*, evidence from clinical research. In what follows it will be shown that this had led to confusion about EBM. A significant problem with these arguments, I shall argue, is not only that they are too imprecise to be helpful, but that the EBM literature is not forthcoming about how to be more precise. In the next section I consider the problems that have arisen interpreting the EBM philosophy of evidence.

5.3 Problems interpreting evidence-based medicine

An outline of the some of the criticisms of EBM is given below. In particular, the role that 'evidence hierarchies' are supposed to play is described.

5.3.1 Criticism of evidence based medicine

It is well documented that EBM has received much criticism, across a wide variety of fronts[325]. In 2000 Sharon Straus and Finlay McAlister undertook a survey of published criticisms of EBM[326]. They found twelve different points, relating to both the practice and the concept of EBM, on which multiple commentators had levelled criticism. In 2004 Aaron Cohen et al categorised and analysed the critical literature about EBM, picking out five key areas of criticism[327]. Also in 2006 Helen Lambert

[325] A small sample of critical literature, over the last decade: (Ashcroft, 2004; H. Brody, F. G. Miller, & Bogdan-Lovis, 2005; A. M. Cohen, Stavri, & Hersh, 2004; Concato, Shah, & Horwitz, 2000; Devisch & S. J. Murray, 2009; Gerber & Lauterbach, 2005; Goldenberg, 2006; Goldenberg, Borgerson, & Bluhm, 2009; Trisha Greenhalgh, 2002; Jenicek, 2006; Lambert, 2006; Loughlin, 2009; M. D. Rawlins, 2008; Straus & McAlister, 2000; R. E. G. Upshur & Colak, 2003; B. A. Williams, 2010; K. Wilson, 2010)
[326] (Straus & McAlister, 2000)
[327] (A. M. Cohen et al., 2004)

produced a list of six limitations to EBM, obtained from a similar survey of the literature[328]. All of these surveys identified the same or similar critical themes, though their categorisations differed. The surveys all picked-out issues clustered around:

(1) EBM denigrates clinical experience.

(2) EBM leads to 'therapeutic nihilism' if there is no evidence from randomised trials.

(3) EBM ignores patients' values and preferences.

(4) EBM is too time-consuming for busy clinicians to practice.

(5) EBM itself lacks good evidence for its own effectiveness.

(3)-(5) relate to the first sense of EBM noted above in §5.1, that is, EBM considered as a an account of evidence for clinical decision making. More important for this discussion are (1) and (2). These are more directly related to the second sense of EBM noted above. That is, EBM considered as an account of evidence for medical claims. These two criticisms each highlight the lack of detail supplied by the basic arguments for EBM. For example: what is the difference between de-emphasising experience and denigrating it? How does one ensure one is doing the former but not the latter? What is one supposed to do if there is no randomised trial evidence? If one lacks evidence from randomised trials, does that mean one cannot have good evidence? If one can have good evidence without a randomised trial, then under what circumstances?

In fact, there is acknowledgement of all these problems in the EBM literature. Most notably, in the 1992 *JAMA* paper[329] , and also in the frequently cited 1996 BMJ paper[330] (both quoted above, in §5.1). Both papers set out to describe ways in which EBM should not be interpreted. However neither paper, nor the literature more broadly, fully addresses these criticisms. They do not alter the view put forward in the arguments rehearsed above in §5.2, rather they both reiterate the non-evidential roles that, for example, clinicians' experience and mechanistic evidence plays. The fact that these criticisms have been acknowledged, but only

[328] (Lambert, 2006) table 1 p. 2634
[329] (Evidence Based Medicine Working Group, 1992)
[330] (D. L. Sackett et al., 1996)

superficially dealt with, further adds to the confusion about the details of EBM in the literature.

Straus and McAlister, as well as Lambert, take an optimistic view about the coherence of EBM, in light of the existing criticisms however. Straus and McAlister label many of the criticisms they identify as misperceptions, misrepresentations or misunderstandings[331]. They state:

> '[such criticisms can] be answered by careful consideration… they represent only pseudolimitations of evidence-based medicine[332]'

The implication here is that these criticisms attack a straw-man and that, in fact, a more sophisticated interpretation of EBM is not susceptible to those criticisms (a similar point to this is also made by Cohen et al[333]). This attitude to the EBM literature assumes that there is a stable notion of what EBM really amounts to – one might say it is 'essentialist'. They respond to criticism with the claim that other commentators have failed to grasp the 'real' nature of EBM. Optimism therefore arises from the view that when the many 'misinterpretations' are ignored, there is a subset of the literature that has indeed correctly captured the EBM philosophy.

Lambert, in a similar vein, argues that the evolution of views about EBM, from the early 1900's to mid-2000's, have been highly accommodating of criticism. She argues that interpretations of EBM have evolved in response to criticism, noting that:

> 'criticism [of EBM] has characteristically been countered not by rejection, contestation or entrenchment, but by incorporation[334]'

This narrative of progress in the EBM philosophy is also seen elsewhere. For example, Adam La Caze claims that views about EBM have 'subtly shifted over time',

[331] A point they have also made elsewhere: (Straus, R Brian Haynes, Glasziou, Dickersin, & G. H. Guyatt, 2007)
[332] (Straus & McAlister, 2000) p. 839 see also: (Straus et al., 2007)
[333] (A. M. Cohen et al., 2004)
[334] (Lambert, 2006) p. 2636

by becoming more sophisticated[335], similarly other recent papers talk in terms of the progress and evolution of EBM[336].

EBM is seen by many as a coherent concept. They hold that some purported criticisms are misplaced, but that genuine criticism has moved the debate on and improved EBM. Other commentators do not assess the EBM literature so positively. Critics have viewed this progress-narrative with cynicism. They argue that, in the face of criticism:

> 'proponents of EBM have continued to 'correct misperceptions' of EBM presumably because to question EBM is surely to misunderstand that which is too obvious to require defence[337]'.

Other critics have also questioned the interpretation of EBM in the literature. In relation to precisely how to fill in the details of the EBM, beyond the basic arguments outlined above, John Worrall has recently claimed that: 'the evidence-based medicine (EBM) movement has got itself into a mess[338]'. His argument can be put in terms of a dilemma: interpretations EBM are either naive to the point where they constitute a view no one would in fact hold, or they are simply unclear about what the interpretation amounts to, because there is no adequately detailed specification of what counts as evidence or how different kinds of evidence are to be balanced. Neither option provides us with a satisfactory interpretation of EBM, therefore EBM is judged to be 'in a mess'.

Brody, Miller and Bogdan-Lovis put forward a similar view[339]. Firstly, they argue that the view taken by Straus et al is broadly correct: critics of EBM have often been guilty of simply misunderstanding and misrepresenting it. Secondly however they argue that the more important area of concern is not with critics' errors, but with the quality of arguments used by EBM's advocates. Like Worrall, they claim that too much of what is said apparently in favour of EBM is unsophisticated and naïve.

I claim that these commentaries on the EBM literature further contribute to the sense that the interpretation of EBM, in so far as it is possible to talk about a

[335] (La Caze, 2008) p. 360
[336] Most notably: (Montori & G. H. Guyatt, 2008)
[337] Beutow 2006 p. 400
[338] (Worrall, 2010) p. 356 See also: (Worrall, 2002, 2007b)
[339] Brody, Miller and Bogdan-Lovis 2005

single interpretation at all, lacks precision. On the one hand there is the essentialist view, which holds that the proper interpretation of EBM has been obscured by misunderstanding and misrepresentation. On the other hand there is a view that holds that the widespread debate indicates a fundamental confusion about the interpretation of EBM (Chapter 6 attempts to distinguish between these views). Importantly, even if there is a coherent account of what the EBM philosophy of evidence *should* be, the claims of widespread misunderstanding or lack of sophistication suggest that is has not been well-articulated by proponents.

One purported remedy to the confusion is EBM's 'evidence hierarchies'. On the face of it they seem to provide a straightforward account of precisely what the EBM philosophy of evidence amounts to.

5.3.2 Evidence Hierarchies and the Categorical Interpretation

Evidence hierarchies rank different kinds of evidence according to the research design employed. In the case where treatment benefit is in question[340], the schema that hierarchies follow is characterised in the following way: (1) clinicians' experience and mechanisms are both placed below controlled clinical research, (2) clinical research is divided into two general kinds: observational studies and, ranking above them, randomised trials, (3) systematic reviews (and, or, meta-analyses) of clinical research are placed above single examples[341]. This hierarchy-schema codifies the idea that one should stress some research designs more than others. Those higher up offer greater evidential support for some purported treatment benefit (Of course, other hierarchy-schema can be devised for addressing treatment harms, or diagnostic test accuracy, etc).

On the face of it evidence hierarchies appear to supply further details on top of the basic arguments for the EBM view. Evidence hierarchies seem to spell out more explicitly what it means to 'de-emphasise' or 'stress' certain kinds of evidence

[340] There are many different hierarchies for ranking evidence for different questions (e.g. treatment harm, diagnostic test accuracy, etc). This is made particularly clear in both of the Oxford Centre for Evidence-Based Medicine's 2009 and 2011 'levels of evidence' tables: (OCEBM Levels of Evidence Working Group, 2009, 2011)

[341] Systematic reviews of RCTs are often shown at the very top of hierarchies, few hierarchies show where systematic reviews of observational studies should be placed (below single RCTs?) One notable exception is: (OCEBM Levels of Evidence Working Group, 2009, 2011) See also (Howick, Chalmers, Glasziou, Trish Greenhalgh, Heneghan, Liberati, Moschetti, Phillips, & Thornton, 2011a)

over others. Indeed, Guyatt and Rennie provide a very simple explanation of how hierarchies can be operationalised:

> 'The hierarchy implies a clear course of action for physicians addressing patient problems: they should look for the highest available evidence from the hierarchy[342]'

As a consequence, a number of authors seem to hold the view that the EBM philosophy of evidence can be read off these hierarchies[343]. That is to say, one can give evidence hierarchies an 'epistemic' reading. Most clearly, Adam La Caze states:

> 'To the extent EBM fills in these philosophical details [of what EBM actually amounts to], it does so by proposing a 'hierarchy of evidence'[344]'

There are of course many examples of hierarchies in the literature, which in various ways add complexity to the schema, above. Hierarchies of evidence can be found in most, if not all, textbooks about EBM[345]. In 2002, a report for the Agency for Healthcare Research and Quality of the US Department of Health and Human Services[346] found 34 different systems for evaluating bodies of clinical evidence[347]; the majority of which[348] relied on some form of hierarchy of research design. Prominent examples of hierarchies include the Oxford Centre for Evidence-Based Medicine (OCEBM) 'levels of evidence #2' table[349], the SIGN system[350], the GRADE system[351], and (pre-dating EBM) the Canadian Task Force on the Periodic Health

[342] (G. H. Guyatt & Drummond Rennie, 2002) p. 13
[343] See for example: (Borgerson, 2009; La Caze, 2008; 2009)
[344] (La Caze, 2008) p. 253-4
[345] (Trisha Greenhalgh, 2006; G. H. Guyatt & Drummond, 2002b; Straus et al., 2005)
[346] (West et al., 2002) p. 45 & 65-6.
[347] And in total, 82 different instruments for rating the quality of particular types of study (namely: Systematic reviews, RCTs, Observational studies, Diagnostic studies) (West et al., 2002) p. 45.
[348] 26/34 See: (West et al., 2002) Grid 5b appendix C pp.134-157
[349] (OCEBM Levels of Evidence Working Group, 2011)
[350] (Harbour, J. Miller, & SIGN Grading Review Group, 2001)
[351] (Grade Working Group, 2004; G. H. Guyatt, Oxman, Kunz, et al., 2008; G. H. Guyatt, Oxman, Vist, et al., 2008)

Examination quality gradings[352]. Notably, the GRADE system and OCEBM levels of evidence #2 table stand out because they also specify conditions under which the evidential support offered by some piece of evidence from a given level may be up- or down-graded.

Frequently the nuances of different hierarchies are glossed over in favour talking about the hierarchy-schema in general. The term 'Categorical Interpretation' has been introduced to characterise one apparently prominent way of understanding the hierarchy-schema above[353].

The Categorical Interpretation holds that a given piece of evidence always gives more evidential support (to the hypothesis in question) than evidence from lower down the hierarchy (which is relevant to the hypothesis[354])[355]. Or put another way, a given piece of evidence is always 'trumped' by evidence above it in the (relevant) hierarchy. So, on the Categorical Interpretation, a randomised trial, for example, will always carry more evidential weight than an observational study investigating the same hypothesis regarding treatment benefit.

A number of authors who have criticised the EBM philosophy of evidence have had something like the Categorical Interpretation in mind as their target. For example, it is not difficult to find statements such as the following:

> 'biologists, astronomers, and chemists would likely be intrigued to learn that certain research methods in medicine are thought to be categorically better than others[356]'

> 'On the Categorical Interpretation, randomisation is seen to provide an incontrovertible epistemic good. The results of randomised studies are epistemologically superior to the results of non-randomised studies, and the superiority is absolute. All the results of a randomised study are always superior to the results of studies from lower down the hierarchy[357]'

[352] (Canadian Task Force On The Periodic Health Examination, 1979) p. 1195
[353] (La Caze, 2008)
[354] Because trivially, a hierarchy of evidence for diagnostic tests is of no use when hypotheses are about, say, treatment benefit.
[355] (La Caze, 2008) p. 354
[356] (Borgerson, 2009) p. 218
[357] (La Caze, 2008) p. 358

> ‘evidence-based medicine has contributed to the development of a rigid hierarchy of research design that underestimates the limitations of randomized, controlled trials, and overstates the limitations of observational studies… [there is a] popular belief that randomized, controlled trials inherently produce gold standard results, and that all observational studies are inferior[358]’

> ‘what these hierarchies claim… is that they [randomised trials] are always better than the alternatives.[359]’

The Categorical Interpretation is often defended as being the dominant interpretation of EBM on the basis of some particularly strong, and therefore often quoted, claims made by EBM’s advocates. In particular the following quotation taken from the definitive EBM textbook *How to Practice and Teach EBM*[360] is held up in support of the Categorical Interpretation:

> ‘If the study wasn’t randomised, we suggest that you stop reading it and go on to the next article in your search[361]’.

More substantial examples of textual support for the Categorical Interpretation include Grossman and McKenzie’s analysis of the *Method for Evaluating Research and Guideline Evidence* (MERGE) document published by the New South Wales Department of Health. They show how the Categorical Interpretation is embedded in the MERGE document[362]. Grossman and McKenzie argue that the MERGE document clearly expresses the view that RCTs are the most rigorous and scientific form of evidence[363]. For instance they argue that the

[358] (Concato, 2004) p. 341 & 346 – see also: (Concato et al., 2000)
[359] Grossman and McKenzie 2005 p. 521
[360] (D. L. Sackett, Straus, W. S. Richardson, Rosenberg, & R Brian Haynes, 2000; Straus et al., 2005)
[361] (Straus et al., 2005) p. 118 – quoted in, for example: (Worrall, 2007b) (La Caze, 2008) (Borgerson, 2009)
[362] (Grossman & Mackenzie, 2005)
[363] (Grossman & Mackenzie, 2005) p. 523

document only makes sense if read under the assumption that it is to be applied to a body of evidence consisting solely of randomised trials[364].

Equally however, we can find examples in the literature which imply a more nuanced interpretation of the EBM philosophy of evidence. One such example comes from the *Users' Guide to the Medical Literature* textbook, where it is stated, simply and briefly that: 'this hierarchy is not absolute[365]'. Other examples include the SIGN[366] and GRADE[367] systems for evaluating evidence, which in contrast to the MERGE document, both explicitly acknowledge that evidence may possess merit (or demerit) beyond its research design. Indeed the GRADE system makes provision for evidence to be upgraded or downgraded according to a number of other quality criteria.

Whether the Categorical Interpretation does truly represent the dominant view in the EBM literature is difficult to assess. As the sample of examples above illustrates, just as there are examples which demonstrate disapproval of the Categorical Interpretation, there are also examples demonstrating approval. The situation is not helped by the fact that, as others have noted, proponents of EBM have not made special attempts to respond to critics, except in the superficial sense noted in §5.3.1[368]. Robyn Bluhm highlights just this problem:

> 'Despite the insistence of the proponents of EBM that, *of course*, no one believes that randomized controlled trials (RCTs) are the only good evidence, or are even *always* the best kind of evidence, it has yet to develop a replacement for a hierarchy that clearly and unequivocally places randomized studies at the top[369]'.

At the very least it is certainly the case that many authors are concerned that EBM is frequently given a Categorical Interpretation[370]; but it may be a worry that those authors with this concern, who claim as a matter of fact that evidence

[364] (Grossman & Mackenzie, 2005) p. 524
[365] (G. H. Guyatt & Drummond Rennie, 2002) p. 13
[366] (Harbour et al., 2001)
[367] (Grade Working Group, 2004; G. H. Guyatt, Oxman, Kunz, et al., 2008; G. H. Guyatt, Oxman, Vist, et al., 2008)
[368] (Buetow, R. Upshur, Miles, & Loughlin, 2006)
[369] (Bluhm, 2010) p. 363 original emphasis.
[370] Consider for example: (Edwards, I. T. Russell, & Stott, 1998)

hierarchies are interpreted categorically, go on to argue that they shouldn't be. This raises the further suspicion they may be building straw men for themselves[371].

5.4 Summary

The basic arguments for EBM put forward a sensible and uncontroversial idea, namely, that medical claims and clinical decisions should be based on the best available evidence. Although the basic arguments for the EBM philosophy of evidence are straightforward, the level of detail they provide is inadequate: when critical attention is turned to the details of EBM, they appear worryingly unclear. The problem was that, while the basic arguments are clear, the conclusions drawn from them were weak. Much extra detail is needed. The task therefore is to give an account of what the EBM philosophy of evidence should amount to. The Categorical Interpretation purports to add clarity and detail, by using evidence hierarchies as its template. However the textual support for the Categorical Interpretation is mixed. Again, it is just not clear whether it truly is the dominant interpretation in the medical literature.

It certainly does seem to be the dominant interpretation in criticisms of homeopathy, however. Debates about homeopathy draw on the resources of EBM to provide an account of 'good evidence' in medicine. As could be seen in the arguments described in Part One, the Canonical Criticism appears to rely on something very close to the Categorical Interpretation. In fact, as we saw in the discussion of the STC Evidence Check report, the justification for evaluating homeopathic treatments in randomised trials was a recapitulation of the basic arguments for EBM, given above. Moreover, proponents of homeopathy accused the Canonical Criticism of being based on a naïve interpretation of EBM that reified evidence from randomised trials. Such a view clearly has much in common with the Categorical Interpretation. If the Categorical Interpretation is dominant, then this lends legitimacy to the arguments put forward by proponents of homeopathy. Furthermore, it is not obvious that opponents of homeopathy have a clearly articulated a more sophisticated interpretation of EBM on which to draw.

[371] (La Caze, 2008) for example acknowledges, then dismisses, the 'wiggle room' which proponents of EBM have to resist the Categorical Interpretation.

In the Chapters which follow therefore, there are both normative and descriptive questions to address: Chapter 6 asks how is EBM interpreted in the literature? More specifically, What interpretations of EBM are there in the literature? How do they relate to the interpretation of EBM, as utilised in the Canonical Criticism? Chapter 7: asks how should EBM be interpreted? More specifically, Is the Categorical Interpretation defensible? What other interpretations of EBM have been proposed?

CHAPTER 6

6. Is there a clear account of evidence-based medicine in the medical literature?

This chapter provides an empirical investigation of the EBM literature. The key questions, from above, are: what interpretations of EBM are there in the literature? How do they relate to the interpretation of EBM, as utilised in the Canonical Criticism? - A number of hypotheses can be formulated:

(1) If, as some claim[372], the Categorical Interpretation is the dominant interpretation in the literature, then one would expect to find that discussion of EBM will be heavily focused on discussions of randomised trials.

(2) If, as others claim[373], the literature contains many 'misperceptions' and 'misrepresentations' of EBM, then one would expect it to be very 'noisy', so there will be:

(2a) many different subsets of papers in the EBM literature, each giving a different interpretation of EBM.

(2b) one subset of the literature (perhaps in the top medical journals, or by prominent advocates of EBM) that represents the 'true' interpretation of EBM.

(3) Given that EBM has been criticised and, so it has been claimed[374], evolved over the past twenty years, one would also expect to find temporal trends in the way that key concepts have been emphasised[375].

This chapter will investigate these hypotheses. The method put forward for this work is essentially a quantitative content analysis of a large set of papers from

[372] See especially: (La Caze, 2008)
[373] See for example: (Straus & McAlister, 2000; Straus et al., 2007)
[374] See for example: (Borgerson, 2009; Buetow, 2009; Montori & G. H. Guyatt, 2008; Worrall, 2007b)
[375] For instance, if one thinks that the interpretation of EBM has evolved to give an increased role patients' values, for example, then there ought to be discoverable trends in the literature which show this.

the medical literature[376]. Multidimensional scaling (MDS) techniques are applied to the results[377]. These methods are further explained below. In §6.1 I describe both the method for defining a set of papers that is representative of the medical literature about EBM, and the method for quantifying the content of the papers in that corpus. In §6.2 the features and characteristics of this data will then be described, along with results from the multidimensional scaling of the corpus. In §6.3 I will evaluate the five hypotheses above, in light of those results.

6.1 Method

6.1.1 Words as data, and multidimensional scaling

The distinctive feature of the text analysis undertaken here is that words and phrases of texts are treated as data[378]. The method involves counting the frequency of key words within a paper and using these frequencies as a measure of that paper's content. A set of papers (the corpus) must be defined and collected for analysis, and a set of key words and phrases (the dictionary) must be specified. When the dictionary is applied to the corpus the result is a quantitative description of each paper. Each paper is expressed as a frequency distribution over the set of key words and phrases, and this provides a measure of the paper's content. This method stands in contrast to approaches that analyse the content of texts through thorough reading, or through the coding of individual sentences or passages; or in general, to approaches where words are understood in context.

The reason for analysing the EBM literature quantitatively is that the texts can be processed electronically. Once a dictionary and a corpus have been defined and the papers in the corpus collected in an appropriate format, the analysis can be automated. In this way, a large number of papers can be analysed in a much shorter time than if they were read and hand-coded. Of course, while one gains in time, one loses some understanding; since however the purpose of this investigation is to

[376] For general introductions to electronic text analysis see, for example: (Adolphs, 2006; Popping, 2000; Weber, 1984)

[377] See below for an explanation. For introductions to multidimensional scaling see, for example: (Borg & Groenen, 1997; T. F. Cox & M. A. A. Cox, 2001; Coxon & Davies, 1982; Davies & Coxon, 1982; Everitt & Rabe-Hesketh, 1997; Kruskal & Wish, 1978)

[378] There are, of course, very many approaches to text analysis in the social sciences, see: (Alexa, 1997)

make claims about the EBM literature as a whole, breadth is more important than depth. The question is whether the literature shows broad trends in the way its key concepts are organised. Consequently the – what one might call – 'low resolution' method of counting key words is an appropriate method.

Multidimensional scaling (MDS) is also used as a way of representing, and aiding the interpretation of, the large data set that the text analysis will generate. MDS encompasses a range of techniques that enables information to be represented geometrically by points in a space. Relationships in the data are reflected in the configuration of the points[379]. Here, the points in the space will represent individual papers in the corpus and the distances between points will reflect the similarities between papers' content, as measured by their scores over the dictionary.

The details of the MDS performed are described in §6.2.3, however an example from Kruskal and Wish usefully illustrates the underlying logic of MDS[380]. Consider that given a map of the UK, one could produce a table showing the distances (in miles, say) between each pair of major cities (such a table would have a diagonal of zeros and be symmetrical about that diagonal, because 'A's distance from B' is a symmetric relation). MDS reverses this process: it provides a method for turning tables of distances back into maps. MDS is useful more generally because the 'distances' involved do not literally need to be distances in terms of meters or miles, nor must the map be confined to only two dimensions. Any data where objects can be characterised in terms of their 'similarity' to each other, lends itself to MDS techniques[381]. MDS allows one to represent those similarities between objects as distances between points. Consequently MDS has been used to investigate a very diverse range of phenomena[382].

The aim of MDS is not merely to re-present the data, but to aid analysis. In this case therefore, MDS is used in order to bring to light the broad conceptual structure of a corpus of papers about EBM; through the interpretation, for example,

[379] (Coxon & Davies, 1982) pp. 1-3 See also: (Kruskal & Wish, 1978)(T. F. Cox & M. A. A. Cox, 2001)

[380] See: (Kruskal & Wish, 1978) pp. 7-9 In fact, this example is a popular one see also: (T. F. Cox & M. A. A. Cox, 2001) p. 2 (Borg & Groenen, 1997) pp. 16-19

[381] See especially: (Everitt & Rabe-Hesketh, 1997)

[382] A particularly eclectic series of examples are discussed in Cox and Cox, covering applications of MDS involving: monkeys, whisky, aeroplanes, yoghurts and bees. Further examples listed but not discussed include: biometrics, counselling psychology, ergonomics, forestry, lexicography, marketing, tourism and brain connectivity. See: (T. F. Cox & M. A. A. Cox, 2001) Ch. 6

of the dimensions along which the points are configured or the interpretation of particular regions in the configuration[383].

The most obvious objection to the method described, which may already be apparent, is that the suggested quantification of the papers' content is too crude to deliver meaningful results. That is to say, the objection amounts to the claim that expressing the content of a paper as a frequency distribution over a pre-specified set of key words and phrases is unlikely to adequately capture the content of that paper[384]. Whilst such a method will certainly not 'fully capture' the content of a paper, it is important to understand that such a goal is not what the method aims at. The adequacy of any method is dependent on the kind of question one wishes to investigate with it. This type of quantitative text analysis is useful when one is interested in prominence and prevalence of various concepts within a set of texts[385]. The occurrence of certain frequently used words or phrases provides an indication of recurring concepts, which can then be studied in terms of whether they do, or do not, occur together with other concepts, and also in terms of whether they relate to further characteristics of the texts. For this purpose, electronic content analysis and MDS have been used effectively, across many fields.

For applying this method to the EBM literature, the three key tasks are: (1) Define a corpus to examine; (2) Build a dictionary; (3) Apply the dictionary and analyse and interpret the findings. These are described below in §6.1.2, §6.1.3 and §6.2-3 respectively.

6.1.2 Define a corpus to examine

The aim is to capture a set of texts that is representative of the literature about EBM. One obvious way this could be done is to stipulate a particular search query to be put through a number of different databases which index medical journals[386]. This is precisely the technique employed by both Straus & McAlister[387]

[383] (Coxon & Davies, 1982) Ch. 4 Describes a number of different, complementary, ways to go about interpreting configurations of points, generated by MDS techniques.
[384] This criticism is not new, see for example: (Goldhammer, 1969; Hays, 1969)
[385] (Popping, 2000) pp. 26 & 42-3. See also: (Alexa, 1997)
[386] For example: PubMed <www.ncbi.nlm.nih.gov/pubmed>, Web Of Knowledge (WOK) <wok.mimas.ac.uk/>, Embase <http://www.embase.com/>, and the Cumulative Index to Nursing and Allied Health Literature (CINAHL) <http://www.ebscohost.com/cinahl/>.
[387] (Straus & McAlister, 2000) – Their search query for the MEDLINE database was: "evidence-based medicine" [MH] OR ("evidence-based" [TW] AND "medicine" [TW]) OR ("evidence"

and by Lambert[388] (discussed in §5.3.1) when they conducted their literature searches. The relevance of the results from search queries are, clearly, determined by the way in which the queries are constructed.

This was not the approach taken here however. Instead the set of potentially relevant texts was defined as those possessing a particular bibliographic property: namely, whether one or more of a smaller set of 'key papers' was cited. The idea behind this approach is that reference to key papers in the field indicates engagement with substantively similar issues. This way of defining a corpus is not contingent on particular search terms, which in this case is a particular advantage. Search terms like "evidence-based medicine" generate an unmanageable number of results. The queries necessary to generate manageable numbers of results require a level of specificity that involves pre-judging characteristics of the papers returned. That is to say, the query used would, in order to generate a manageable number of results, need to be specified to a degree which required substantive commitment to the kind of results one wanted to see returned[389].

Consequently such a method would be in danger of begging the question, since the aim is to discover characteristics of the literature. Another advantage of defining the corpus using a bibliographic property is that it is more likely that a contribution to a topic will cite the key literature in that topic, than include key words in its title[390]. I claim there is a prima facie reason to think that the bibliographic method is superior to the search-query method for defining a corpus of papers about a given topic; at least in cases where there is a set of acknowledged key papers to refer to. As I will explain below, this is the case for EBM.

One constraint imposed by this approach is that it restricts the corpus to papers published in journals indexed in the WOK database; since it is through WOK that the citation data are available (in an efficiently accessible way). Consequently it excludes papers published in journals which are not indexed, and also excludes guidebooks, textbooks, and book chapters. It should be noted that there are some

---

[TW] and "based" [TW] and "medicine" [TW]) AND "limitations" [MH] OR "criticisms" [MH] OR "limitations" [TW] OR "criticisms" [TW].

[388] (Lambert, 2006) – Search query for WOK was: 'EBM' AND 'crit*' OR 'limit*'

[389] As can be seen for example by the inclusion of 'limitation' and 'criticism' in the queries cited in footnotes above.

[390] One example of an article which does precisely that is (Tanenbaum, 1993). It does not appear in the WOK results for the search 'evidence-based medicine', but does cite the original 1992 JAMA paper and is, uncontroversially, about EBM.

arguably important texts about EBM which are excluded from the corpus – for instance the multiple editions of the definitive EBM textbook *How to Practice and Teach EBM*[391], or the collection of critical essays *Evidence-Based Medicine: In its place*[392].

As compared to these other sources, an analysis of the papers in the WOK database offers a large and more varied body of literature to make claims about. Moreover it might be argued that defining the corpus only in terms of the papers in the WOK database are sufficient to capture any arguments or opinions that might also occur in other sources.

The key papers, from which the corpus is constructed, were determined as follows: A preliminary list of key papers (based on personal knowledge of the literature) was generated. Even a rough knowledge of the development of EBM is sufficient to suggest obvious candidate papers – such as the 1992 JAMA paper[393] and the 1996[394] BMJ paper repeatedly mentioned above.

In order to validate this list, and to add further papers, two 'expert surveys' were conducted. Members of one (or both) of two internet mailing-lists: the Evidence-Based-Health mailing-list[395] and a Philosophy of Medicine mailing-list were surveyed. The Evidence-Based-Health mailing-list is organised by the Oxford Centre for Evidence Based Medicine (CEBM)[396] and it was through the CEBM website that the mailing-list was found. The philosophy of medicine mailing-list is administered by the International Philosophy of Medicine Roundtable[397] and consists of philosophers and clinicians whose work is relevant to topics in the philosophy of medicine.

In both cases the same message was sent to the mailing-list members. The message explained the aim of producing a list of key papers in the EBM literature and presented seven papers as possible examples (see table below). Mailing-list members were asked to contribute their list of candidate papers.

From the Evidence-Based-Health mailing list, the response was poor: the response was better from the philosophy of medicine list. In total twelve responses

[391] (D. L. Sackett et al., 2000) (Straus et al., 2005)
[392] (Kristiansen & Mooney, 2004)
[393] (Evidence Based Medicine Working Group, 1992)
[394] (D. L. Sackett et al., 1996)
[395] <http://www.jiscmail.ac.uk/lists/EVIDENCE-BASED-HEALTH.html>
[396] <http://www.cebm.net/>
[397] <https://sites.google.com/site/philosmed/>

were received, with much overlap between the responses. This resulted in the following list of 'key papers' (those marked with an * come from the survey):

TABLE 6.1: Key EBM Papers

Papers	Number of times cited in WOK database (up to 28/03/2010)	From Surveys
Evidence Based Medicine Working Group 1992: 'Evidence Based Medicine: A New Approach to Teaching the Practice of Medicine'. *JAMA*, 268, pp. 2420-25.	915	
Tanenbaum, S.J. 1993: 'What Physicians Know'. *New England Journal of Medicine*, 329, pp. 1268-71.	130	*
Feinstein, AR 1994: 'Clinical judgment revisited: the distraction of quantitative models'. *Annals of Internal Medicine*, 120, p. 799.	92	*
Naylor, C. D. 1995: 'Grey zones of clinical practice: some limits to evidence-based medicine'. *Lancet*, 345, pp. 840-42.	325	*
Rosenberg, W.M.C. and Donald, A. 1995: 'Evidence based medicine: an approach to clinical problem solving'. *BMJ*, 310, pp. 1122-6.	341	
Sackett, D., Rosenberg, W.M.C., Gray, J. A. Muir, Haynes, R. Brian and Richardson, W. S. 1996: 'Evidence based medicine: what it is and what it isn't'. *BMJ*, 312, pp. 71-2.	2040	
Feinstein, A. and Horwitz, R. I. 1997: 'Problems in the "Evidence" of "Evidence-Based Medicine"'. *The American Journal of Medicine*, 103, pp. 529-35.	237	*
Charlton, B. G. and Miles, A. 1998: 'The rise and fall of EBM'. *QJM*, 91, pp. 371-74.	50	*
Tonelli, MR 1998: 'The philosophical limits of evidence-based medicine'. *Academic Medicine*, 73, p. 1234.	82	*
Straus, Sharon E. and McAlister, Finlay A. 2000: 'Evidence-based Medicine: a commentary on common criticisms'. *Canadian Medical Association Journal*, 163, pp. 837-41.	103	
Rodwin, MA 2001: 'The politics of evidence-based medicine'. *Journal of health politics, policy and law*, 26, p. 439.	13	*
Norman, GR 2001: 'Examining the assumptions of evidence-based medicine'. *Journal of Evaluation in Clinical Practice*, 5, pp. 139-47.	33	*
Worrall, J. 2002: 'What Evidence in Evidence-Based Medicine'. *Philosophy of Science*, 69, pp. 316-30.	17	*
Haynes, RB 2002: 'What kind of evidence is it that Evidence-Based Medicine advocates want health care providers and consumers to pay attention to?'. *BMC Health Services Research*, 2, p. 3.	61	*
Gupta, M 2003: 'A critical appraisal of evidence-based medicine: some ethical considerations'. *Journal of Evaluation in Clinical Practice*, 9, pp. 111-21.	33	*
Sehon, SR and Stanley, DE 2003: 'A philosophical analysis of the evidence-based medicine debate'. *BMC Health Services Research*, 3, p. 14.	20	*
Ashcroft, R 2004: 'Current epistemological problems in evidence-based medicine'. *Journal of Medical Ethics*, 30, pp. 131-35.	17	*
Cohen, AM, Stavri, PZ and Hersh, WR 2004: 'A categorization and analysis of the criticisms of evidence-based medicine'. *International journal of medical informatics*, 73, pp. 35-43.	48	*

Eddy, D.M. 2005: 'Evidence-based Medicine: A Unified Approach'. *Health Affairs*, 24, pp. 9-17.	52	
Tonelli, MR 2006: 'Integrating evidence into clinical practice: an alternative to evidence-based approaches'. *Journal of Evaluation in Clinical Practice*, 12, pp. 248-56.	38	*
Montori, V.M. and Guyatt, G. 2008: 'Progress in Evidence-Based Medicine'. *JAMA*, 300, pp. 1814-6.	11	
Djulbegovic, B, Guyatt, GH and Ashcroft, RE 2009: 'Epistemologic inquiries in evidence-based medicine'. *Cancer Control*, 16, pp. 158-68.	11	
Total Citations	4669	
Duplicates	1170	
Unique Papers Citing	3499	

The corpus was generated by collecting the details of all papers indexed in the WOK database citing one or more of the key papers. As shown in the table above, this amounted to 3,495 unique papers[398].

There are two reasons for reducing the size of this corpus further. Firstly, because 3,500 papers are too many to analyse: even electronic analysis requires that papers be 'cleaned up' and converted into the appropriate format and while this can be partially automated 3,500 papers was judged to require too much time to process. Secondly and more importantly, those 3,500 papers were not all directly relevant.

A broad categorisation of the corpus was undertaken in order to determine the different ways in which they were 'about' EBM. Clearly all the papers in the corpus had some connection to EBM, simply in virtue of the fact that they cited one or more of the key EBM papers; however there were no obvious prior grounds on which to distinguish them. A random sample of 200 papers was selected in order to generate a set of categories that could be applied to the corpus. Papers falling under the least salient categories could then be discarded from the corpus.

The categorisation process involved determining, from the content of the abstracts or titles of papers, the different ways in which they may, or may not,

[398] Because many of the papers in the corpus cite multiple key papers it turns out that the full list of 23 key papers was unnecessary in generating the vast majority of the 3,495. Had the references been collected by going through the key papers sequentially, in order of most to least cited, the number of unique corpus papers contributed by the marginal key paper, while never zero, quickly diminished. Thus it would have been possible to generate a substantially similar corpus from fewer key papers. Removing duplicates sequentially is, however, more time consuming than removing them in one go. This is why the sequential approach was not adopted.

directly engage with EBM[399]. It is worth noting that, from the sample of 200 papers, it became obvious that some only included a 'throw-away' citation to one of the key papers, that is to say, they cited a key paper in a merely incidental way. Hence a clear way to reduce the size of the corpus was to identify and remove papers of this sort. Other papers in the sample of 200 discussed issues that are only tangentially related to EBM. In most cases, these were papers situated in healthcare disciplines other than medicine, to which the prefix 'evidence-based' had been applied. So for example, very many papers were about evidence-based nursing, surgery, dentistry, radiology, physical therapy, social work, management, or health policy; or 'evidence-based practice' which served as a term that could denote any and all of the above, as well as medicine. Consequently papers about these topics were removed from the corpus, unless they were judged to contain discussion of EBM too. For example a paper purely about evidence-based surgery would have been excluded, whereas a paper about the relationship between evidence-based medicine and evidence-based nursing would not. Additionally, there were numerous papers that reported empirical work, for instance investigating attitudes towards EBM among various populations such as GPs, nurses, patient groups etc. These too were excluded.

It is worth noting also that, as well as there being many irrelevant papers there are some important omissions. For example the series of papers published in the *BMJ* in 2008, about the GRADE framework for evaluating evidence do not cite any of the key papers that were used to generate the corpus. Hence, they do not feature in the corpus. This is not necessarily problematic however, the corpus must be representative of the EBM literature but there is no requirement for it to be exhaustive. The purpose is not to define a corpus which *is* 'the EBM literature'. Indeed given the way the corpus was reduced in size it is unlikely that, were the process repeated by another individual, the very same set of papers would emerge. I do claim however that a very similar set would emerge – the corpus is robust in that sense. The purpose of defining the corpus in this way is to produce a large set of papers that is representative of the EBM literature. I claim that the corpus is an excellent basis for inferences about the EBM literature as a whole, even if it does not comprehensively capture it and even if it is (trivially) sensitive to replication. It is

[399] The title alone was not judged to be sufficiently reliable to categorise the papers, even taking into account the fact it would have made the categorisation process quicker – though obviously where abstracts were not available only the titles could be used (33% [1,182] of the 3,500)

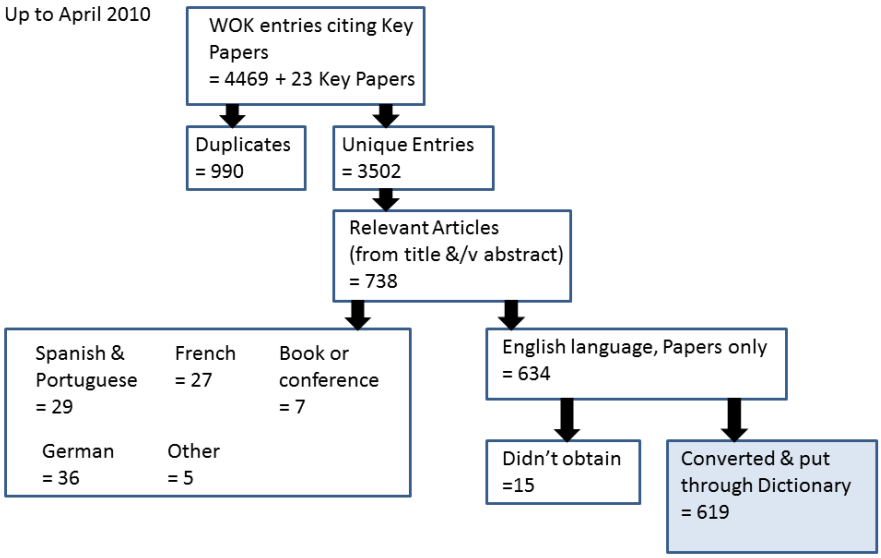
perhaps disappointing that the GRADE papers do not feature, but this is unlikely to make a substantial difference to the analysis: the size of the (pared down) corpus compensates for these individual omissions.

From the sample of 200 papers, the following 6 non-exclusive categories emerged as an adequate way to categorise the papers:

(1) **Scope, Limits & Methods of EBM - logic of EBM**
(2) **Kinds or Nature of Evidence, in General**
(3) Information Management & Critical Appraisal - How to do EBM
(4) Clinical Practice & Guidelines - Using EBM
(5) **Histories or Overviews of EBM**
(6) Medical Education and EBM

All 3,500 papers in the corpus were assigned to one or more of these six categories. Only three of these categories pick out papers of interest (shown in bold). 731 papers in the corpus fell under at least one of the three categories[400]. The process is shown in the diagram below:

FIGURE 6.2: Generating and refining the corpus



[400] References for these papers can be found in the data tables. They are not included in the bibliography of the thesis.

6.1.3 Building a dictionary

It is crucial for the adequacy of this type of content analysis that the concepts used 'span the meaning space of the texts[401]' and that the individual words and phrases are assigned to specific concepts with high validity[402]. One can begin to meet both of these points by constructing the 'dictionary' – that is, the set words to be counted – specifically for the research in question; which in this case means generating 'inductively'[403] the sets of relevant key words and phrases that will cover texts about EBM[404].

The process for constructing the dictionary involves taking a representative sample of papers in the corpus and extracting the most frequently-used and substantively interesting words or phrases. These constitute the dictionary entries and are the words and phrases that will be counted in each of the corpus papers. The dictionary entries are also categorised thematically, in order to identify the concepts that are most prominent in the corpus.

A stratified sample of papers was taken. Any papers falling under (1), (2) and (5) above, which were judged to be paradigm examples of each category, were divided into groups according to their publication date. One paper from each category and date-group was randomly selected, as shown in the following table:

TABLE 6.3: Stratified sample of papers used to construct the dictionary

Relevant Categories	1994-6	1997-9	2000-2	2003-5	2006-10
Scope & Limits of EBM	#3828	#2906	#1594	#1065	#4456
Evidence, generally	#3821	#1873	#4071	#875	#692
History or Overview	#3886	#1728	#4074	#3035	#2194

Id #	Reference
692	Tonelli, M.R., 2006. Integrating evidence into clinical practice: an alternative to evidence-based approaches. *Journal of Evaluation in Clinical Practice*, 12(3), pp.248-256.

[401] (Popping, 2000) p. 45
[402] (Popping, 2000) pp. 45-6
[403] As opposed to 'deductively', where the key words and concepts are pre-specified on theoretical grounds, or come from a 'general dictionary'. See: (Popping, 2000) p. 45 (Weber, 1984) p. 133
[404] (Popping, 2000) pp. 40-59 (Alexa, 1997)

875	Ernst, E., 2005. On the inconclusiveness of "evidence." *Wiener klinische Wochenschrift*, 117(7-8), pp.241-242.
1065	Cohen, A.M., Stavri, P.Z. & Hersh, W.R., 2004. A categorization and analysis of the criticisms of Evidence-Based Medicine. *International Journal of Medical Informatics*, 73(1), pp.35-43.
1594	Straus, S.E. & McAlister, F.A., 2000. Evidence-based medicine: a commentary on common criticisms. *Canadian Medical Association Journal*, 163(7), pp.837-841.
1728	Antes, G., Galandi, D. & Bouillon, B., 1999. What is evidence-based medicine? *Langenbeck's Archives of Surgery*, 384(5), pp.409-416.
1873	Edwards, A.G.K., Russell, I.T. & Stott, N.C.H., 1998. Signal versus noise in the evidence base for medicine: an alternative to hierarchies of evidence? *Family Practice*, 15(4), pp.319-322.
2194	Darzi, A., 2008. Evidence-based medicine and the NHS: a commentary. *Journal of the Royal Society of Medicine*, 101(7), pp.342-4.
2906	Charlton, B.G. & Miles, A., 1998. The rise and fall of EBM. *QJM*, 91(5), pp.371-374.
3035	Akobeng, A.K., 2005. Principles of evidence based medicine. *Archives of Disease in Childhood*, 90(8), pp.837-40.
3821	Feinstein, A.R., 1995. Meta-analysis: Statistical Alchemy for the 21st Century. *Journal of Clinical Epidemiology*, 48(1), pp.71-79.
3828	Naylor, C.D., 1995. Grey zones of clinical practice: some limits to evidence-based medicine. *Lancet*, 345, pp.840-842.
3886	Vandenbroucke, J.P., 1996. Evidence-Based Medicine and "Medecine d'Observation." *Journal of Clinical Epidemiology*, 49(12), pp.1335-1338.
4071	Upshur, R.E.G., 2000. Seven characteristics of medical evidence. *Journal of Evaluation in Clinical Practice*, 6(2), pp.93-97.
4074	Wolf, F.M., 2000. Lessons to be learned from evidence-based medicine: practice and promise of evidence-based medicine and evidence-based education. *Medical Teacher*, 22(3), pp.251-259.
4456	La Caze, A., 2009. Evidence-Based Medicine Must Be …. *Journal of Medicine and Philosophy*, 34(5), pp.509-527.

The full-texts of the papers listed in the table above were obtained, and these were converted into plain-text files, containing only the text from the body of the article; with numbers punctuation etc removed. Additionally the redundant words of the text were removed: words such as "and", "it", "highly", "almost" etc[405]. From this, frequency and concordance[406] reports were generated for the remaining words. Approximately 250 key words were identified that are substantively related to EBM.

Note that the list of key words also included phrases, such as (most obviously) 'evidence-based medicine'. There was some difficulty in determining the most suitable phrases to include in the dictionary. For example, the word 'clinical' is one of

[405] See: (Popping, 2000) p. 41
[406] Sometime called KWIC (Key Word In Context) reports.

the most frequently occurring words, and it is often found to form adjectival phrases with many other words (such as 'practice', 'experience', or 'research'); for precisely this reason however it is almost entirely redundant when taken in isolation, because it has no stable meaning. In other cases it was difficult to determine the appropriate length of phrase to include in the dictionary. Consider, that 'best available evidence' is one such frequently occurring phrase. Examination of concordance reports suggested that 'best' could be singled out individually, as it was rarely used in any other context than this or equivalent phrases, whereas 'available' was used in many other contexts and, like 'clinical', was for that reason unhelpful for the dictionary.

The key words and phrases were grouped semantically; checking against concordance reports that such grouping was valid[407]. Furthermore words were lemmatised[408] at this point, again with reference to the concordance reports. Once this semantic grouping was complete the words and phrases, or groups of words and phrases, were further categorised thematically. As before this process involved checking the categorisations against the concordance reports to ensure that they were genuinely similar in meaning. Consequently, a list of dictionary categories was generated each made up by a number of related dictionary entries. The list of the dictionary categories is given below; for the complete dictionary see the corresponding data file noted at the beginning of §6.2.

[407] See: (Popping, 2000) p. 43

[408] For example, 'evidence', 'evidential', 'evident' can be lemmatised to 'eviden*'.

TABLE 6.4 Dictionary Categories[409]

Dictionary Category Name	Description:
Against & For	Counts words such as debate, proponents and critics
CAM	Counts words such as placebo and homeopathy
Context	Counts words such as context, social and bedside
Criticism	Counts words that refer to different kinds of criticism of EBM
Dealing with Evidence	Counts words that refer to the appraisal, quality and weighing of evidence
EBM	Counts occurrences of 'evidence-based medicine' and 'EBM'
Effectiveness	Counts words that refer to benefits, effectiveness or ineffectiveness.
Fair Test Concepts	Counts that refer to randomisation, blinding, bias and control
Important	Counts words such as important, crucial and emphasis
Kinds of Evidence	Counts words such as scientific evidence, medical evidence, empirical evidence
Kinds of Experiment	Counts words that refer to different kinds of clinical experiments
Knowledge, Experience & Skills	Counts words that refer to judgement, knowledge and practice
Methods	Counts words such as design, compare, approach etc.
Patients	Counts words that refer to people or patients
Philosophy	Counts words that refer to philosophical theories
Preferences	Counts words such as choice, value and preferences
Principles	Counts words such as concept, idea and principle
Problematic Kinds of Evidence	Counts words that refer to mechanistic, physiological and problematic evidence
Professionals	Counts words that refer to clinical and non-clinical healthcare professionals
Treatments	Counts words such as treatment, therapy, care
Views	Counts words such as view, dogma, paradigm

The dictionary entries were inputted into the free-software *The Yoshikoder*[410], along with a plain-text version of each paper in the corpus. *The Yoshikoder* then outputs a spread sheet where the rows list the individual corpus papers and the

[409] Note that while I have tried to name the categories to reasonably represent the words they count, the meaning of the categories becomes more apparent when the dictionary is consulted directly.

[410] <http://www.yoshikoder.org/> - 'The Yoshikoder is a cross-platform multilingual content analysis program developed as part of the Identity Project at Harvard's Weatherhead Center for International Affairs.'

columns list the dictionary entries. Hence, each row shows a single paper's score for each of the dictionary entries[411]. This was the starting point for the analysis. The software *PASW Statistics 18*[412] was used throughout. The raw dictionary scores for each paper were converted to proportions of the paper's total score over the dictionary. Longer papers would naturally be expected to have higher raw scores for each dictionary entry simply in virtue of their greater length. Hence the purpose of expressing the frequencies as proportions of the total number of key words mentioned in a given paper is to control for the length of the paper. Consider that what is significant is not that a given paper uses more key words than another, but that the percentage of certain kinds of key words are different between papers; since it is this that is likely to indicate a different emphasis and focus. Consequently the MDS was performed on each papers dictionary scores expressed as a proportion of the total number of key words counted in that paper. The next section describes the results of the analysis of this data.

6.2 Results

Data files

(1) EBM corpus data spread sheet[413]: <http://goo.gl/IqAh7>

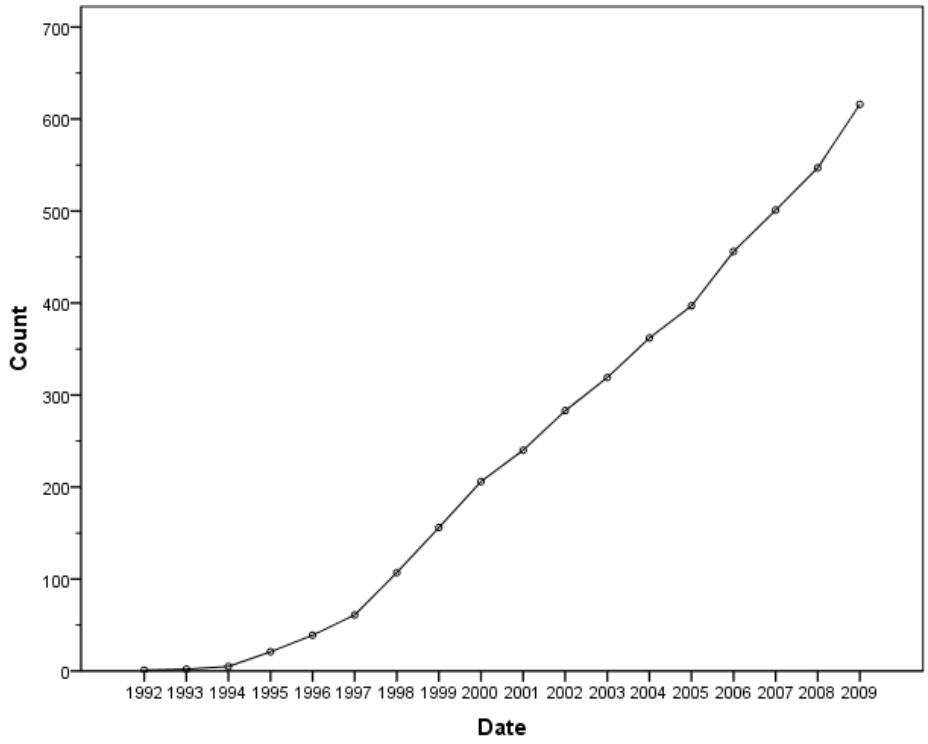
(2) Dictionary spread sheet: <http://goo.gl/6oLbw>

6.2.1 Characteristics of the corpus: journals and authors

The corpus is made up of 619 papers, including the 23 key EBM papers that were used to generate the corpus. From 1994 there is a continuous and almost constant rate of growth of the corpus:

[411] See below for links to data files.
[412] <http://www-01.ibm.com/software/analytics/spss/>
[413] Hard copies are available on request (andrewjamesturner0@gmail.com)

FIGURE 6.5: Cumulative growth of papers in the corpus[414]



[414] 2010 has been removed, because there was not a full years' data at the time of collection (April 2010)

Figure 6.5 shows that the corpus has grown by approximately the same amount each year, since 1994. After nearly twenty years of EBM being a named concept in the medical literature, the number of papers engaging explicitly with EBM has not plateaued. This is especially interesting given the level of acceptance that EBM has achieved over that time: one might reasonably expect there to be less new literature which explicitly engages with the concept. As can be seen from figure 6.5 however, this is not the case.

The papers in the corpus are published in 305 different journals: 180 (29%) of the papers in the corpus are published in the 11 most-published-in journals, and 315 (51%) in the top 50. The 11 most-published-in journals are as follows:

TABLE 6.6: Top eleven most-published-in journals in the corpus

Journal	Count
Journal of Evaluation in Clinical Practice (JECP)	68
British Medical Journal (BMJ)	30
Lancet	13
Perspectives in Biology and Medicine	13
Annals of Internal Medicine	10
Social Science & Medicine	9
Academic Medicine	8
British Journal of General Practice	8
Journal of the American Medical Association (JAMA)	7
Journal of Clinical Epidemiology	7
Theoretical Medicine & Bioethics	7

The large number of papers from the Journal of Evaluation in Clinical Practice is due to their special issues dedicated to EBM[415]. Similarly, the less well known journal[416] Perspectives in Biology and Medicine also features as one of the most published in journals on account of its special issues on EBM[417]. The presence of four of the five highest impact[418] general medical journals (BMJ, Lancet, JAMA, Annals of Internal Medicine) confirms that debates about EBM are prominent and considered

[415] In particular, the corpus contains many papers from volumes 9, 12, 14 & 15.
[416] In comparison, that is, to the other journals above and beneath it in table 6.6.
[417] Volumes 48 & 52.
[418] Based on impact factors in the "Medicine – General & Internal" category of the Thomson Reuters 2009 Journal Citation Report:
<http://admin-apps.isiknowledge.com/JCR/JCR?RQ=HOME>

important, although it is notable that the New England Journal of Medicine does not feature in the table[419]. UK, US and Canadian journals dominate the corpus. 247 (40%) of papers in the corpus were published in UK journals and 224 (36%) were published in US or Canadian journals, where EBM and clinical epidemiology have their intellectual roots. However, it is important to note that non-English language papers were excluded from the corpus, and so this characteristic of the corpus is unsurprising[420].

The authors who appear most often in the corpus are shown in the table below:

TABLE 6.7: Top thirteen most published authors in the corpus

Author	Count[1]
Miles *	12
Upshur *	11
Haynes	9
Loughlin *	9
Guyatt	8
Sackett	7
Tonelli *	7
Buetow *	6
Charlton *	6
Polychronis *	6
Cook	5
Feinstein	5
Wyer	5

[1] Some co-authorship between these authors means that the sum of the count column exceeds the number of unique papers (74)

* More 'critical' authors

Many of the most published authors in the corpus are also authors of the key papers that were used to generate the corpus. While this might be thought to reflect a bias towards self-citation, I suggest it is more likely to reflect the fact that these authors genuinely are 'key players' in the EBM literature. Notice also that there is an

[419] There are only two NEJM papers in the corpus.

[420] Moreover there is an English language bias in the WOK index, as noted by Thomson Reuters on their Journal Selection Process web-page: <http://wokinfo.com/benefits/essays/journalselection/>

almost even split between authors that – in a very broad sense – either advocate or criticise EBM.

6.2.2 Analysis of the corpus

6.2.2.1 Correlations of publication date and journal impact factor with the dictionary categories

With respect to the whole corpus the structure of emphasis over the dictionary categories does not vary significantly according to either the date a paper was published or the impact factor of the journal it was published in. This pattern changes slightly when one considers only the most published authors in the corpus. In that case, there is a group of three dictionary categories ("Against and For" (r=.558[421]), "Evidence Based Medicine" (r=.588), and "Philosophy" (r=.544)) which score higher over time, and a group of four dictionary categories ("Fair Test Concepts" (r=-.329), "Kinds of Experiment" (r=-.491), "Effectiveness" (r=-.353) and "Treatments" (r=-.379)) which score lower over time. It is also the case that authors advocating EBM score increasingly well with time on the "preferences" category (r=.524). The full table of correlations is presented below:

[421] Here, and in what follows, values are only quoted if they are significant at the 0.01 level (unless stated otherwise).

TABLE 6.8: Correlations of Date and Impact Factor with the Dictionary Categories, for the corpus and subgroups (significant correlations highlighted)

			AGAINST AND FOR	CAM	CRITICISM	DEALING WITH EVIDENCE	EVIDENCE BASED MEDICINE	IMPORTANT	KINDS OF EVIDENCE	KNOWLEDGE, EXPERIENCE & SKILLS	FAIR TEST CONCEPTS	KINDS OF EXPERIMENT	METHODS	PATIENTS	PREFERENCES	PHILOSOPHY	PRINCIPLES	PROBLEMATIC KINDS OF EVIDENCE	PROFESSIONALS	CONTEXT	EFFECTIVENESS	TREATMENTS	VIEWS
Corpus	Date	Pearson Correlation	.113**	0.038	0.031	-0.054	.132**	-0.046	-0.037	-0.021	0.06	-0.037	0.051	-0.047	.087*	.186**	.087*	0.042	-.098*	0.047	-0.05	-.081*	0.052
		N	619	619	619	619	619	619	619	619	619	619	619	619	619	619	619	619	619	619	619	619	619
	Impact factor	Pearson Correlation	-.088*	-0.01	-0.043	0.024	-0.07	0.016	-0.002	-0.032	-0.016	0.069	-0.058	0.035	0.017	-0.076	-0.03	-0.035	0.016	-.085*	.094*	0.022	-0.066
		N	602	602	602	602	602	602	602	602	602	602	602	602	602	602	602	602	602	602	602	602	602
Key Papers	Date	Pearson Correlation	0.275	0.343	-0.121	0.136	0.377	0.309	-0.17	-0.096	-0.111	-0.188	0.152	-0.302	0.393	.451*	0.36	0.184	-.432*	.569**	-0.278	-0.2	-0.047
		N	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22
	Impact factor	Pearson Correlation	-.467*	-0.242	-0.162	-0.025	-.433*	-0.024	0.164	0.268	0.142	-0.171	-0.236	-0.076	0.094	-0.111	-0.17	-0.09	0.405	-0.264	.573**	0.15	-0.042
		N	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21
Papers by key Proponents	Date	Pearson Correlation	.410*	-0.062	-0.176	-0.101	.376*	0.077	-0.077	.379*	-0.254	-0.256	-0.327	0.099	.524**	.475**	-0.15	0.072	-0.024	0.036	-0.097	-0.339	-0.119
		N	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31
	Impact factor	Pearson Correlation	-0.248	-0.169	-0.23	.399*	0.104	-0.11	-0.041	0.142	-0.171	-0.098	-0.231	-0.081	0.217	-0.202	0.33	-0.057	-0.015	0.351	0.046	-0.185	-0.131
		N	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).
Proponents = Cook, Feinstein, Guyatt, Haynes, Sackett, Wyer

6.2.2.2 Correlations between the dictionary categories

Correlations between the dictionary categories shows the extent to which, if one dictionary category scores low or high, other categories score low or high with it. Notably, few correlations between dictionary categories fall within the set {r: |r| > .3} and all correlations are within {r: |r| < .51}. Most dictionary categories therefore are largely independent of each other. Never the less, those correlations that fall within {r: |r| > .3} are worth noting:

Firstly, “Fair test concepts” and “kinds of experiments” are positively correlated with each other (r=.405), but both are negatively correlated with “dealing with evidence” (r=-.292 and -.157, respectively) and with “evidence based medicine” (r=-.279 and -.372, respectively). Secondly, “knowledge, experience & skills” is positively correlated with “professionals” (r= .326) and with “context” (r= .303), as well as negatively correlated with “fair test concepts” (r= -.301), “kinds of experiment” (r= -.509), “effectiveness” (r= -.325) and “treatments” (r= -.309)[422]. Thirdly, “against and for” is positively correlated with “criticism” (r= .389), “evidence based medicine” (r= .395) and “philosophy” (r= .397). The full table of correlations is given below:

[422] “effectiveness” and “treatments” are themselves moderately correlated (r=.489); most likely because talk of effectiveness often occurs in the context of the effectiveness *of treatments*.

TABLE 6.9: Correlations Between Dictionary Categories: |r| > .3 highlighted

		CAM	CRITICISM	DEALING WITH EVIDENCE	EVIDENCE BASED MEDICINE	IMPORTANT	KINDS OF EVIDENCE	KNOWLEDGE, EXPERIENCE & SKILLS	FAIR TEST CONCEPTS	KINDS OF EXPERIMENT	METHODS	PATIENTS	PREFERENCES	PHILOSOPHY	PRINCIPLES	PROBLEMATIC KINDS OF EVIDENCE	PROFESSIONALS	CONTEXT	EFFECTIVENESS	TREATMENTS	VIEWS
AGAINST AND FOR	Pearson Correlation	-0.07	.389**	-.140**	.395**	0.016	-.112**	.115**	-.096*	-.225**	-0.05	-.207**	-0.03	.397**	.081*	.124**	-0.02	.116**	-.241**	-.254**	.320**
CAM	Pearson Correlation		-.102*	-.143**	-.123**	-0	-.100*	-.155**	0.051	.142**	0.075	0.077	-0.06	-0.06	-0.02	-0.043	-0.06	-.099*	.149**	.152**	-0.05
CRITICISM	Pearson Correlation			-0.08	.210**	0.002	-0.07	0.046	-.096*	-.185**	0.026	-.195**	-.131**	.199**	0.067	0.023	-0.05	0.037	-.138**	-.194**	.167**
DEALING WITH EVIDENCE	Pearson Correlation				.166**	-0.05	.093*	-0.031	-.292**	-.157**	-.153**	-.352**	0.041	-0.06	0.031	-.104**	-.130**	-0.06	-.199**	-.218**	-0.05
EVIDENCE BASED MEDICINE	Pearson Correlation					-.149**	-.173**	0.078	-.279**	-.372**	-.203**	-.255**	-0.02	.337**	.117**	0.02	0.072	-0.01	-.298**	-.314**	.262**
IMPORTANT	Pearson Correlation						0.055	0.005	-.095*	-0	-0.01	0.046	-0	-0.04	-0.05	-0.001	-0.05	.125**	-0.03	0.005	0
KINDS OF EVIDENCE	Pearson Correlation							-0.055	-0.06	0.055	0.011	.093*	-0.03	-.114**	-.100*	-0.043	-0.07	-0.06	-.091*	-.102*	-.133**
KNOWLEDGE, EXPERIENCE & SKILLS	Pearson Correlation								-.301**	-.509**	-.243**	-.080*	.121**	.203**	0.051	.086*	.326**	.303**	-.325**	-.309**	.142**
FAIR TEST CONCEPTS	Pearson Correlation									.405**	.250**	-.083*	-.086*	-0.06	-0.02	-0.019	-.208**	-.140**	0.065	-0.06	-.123**
KINDS OF EXPERIMENT	Pearson Correlation										.149**	-.181**	-.223**	-.284**	-.172**	-.087*	-.379**	-.285**	.111**	0.056	-.200**
METHODS	Pearson Correlation											-.104**	-.109**	0.002	0.007	-0.028	-.220**	0.06	-0.02	-0.02	0.02
PATIENTS	Pearson Correlation												.139**	-.180**	-.102*	-0.027	.168**	-0.02	.190**	.241**	-.129**
PREFERENCES	Pearson Correlation													0	.117**	-0.056	-0.01	.116**	-0.08	-0.07	-0.02
PHILOSOPHY	Pearson Correlation														.193**	.170**	-0.01	.133**	-.253**	-.276**	.255**
PRINCIPLES	Pearson Correlation															0.07	0.031	0.07	-.167**	-.165**	.168**
PROBLEMATIC KINDS OF EVIDENCE	Pearson Correlation																-0.01	0.032	-.112**	-0.07	.160**
PROFESSIONALS	Pearson Correlation																	.107**	-.136**	-.087*	0.053
CONTEXT	Pearson Correlation																		-.212**	-.094*	0.066
EFFECTIVENESS	Pearson Correlation																			.489**	-.203**
TREATMENTS	Pearson Correlation																				-.215**

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

The correlation table, and three examples picked out above, show that there is at least some further structure to the way that EBM is talked about. Multidimensional scaling (MDS) helps us to explore this structure further[423].

6.2.3 Multidimensional Scaling

Multidimensional scaling was used to plot papers in the corpus as points in a space, where the distance between points represents a measure of dissimilarity between papers, calculated from each paper's score over the dictionary.

A matrix of proximities was produced by calculating from the dictionary scores the Euclidean distance between pairs of papers[424]; that is to say, the dissimilarity between papers was calculated as the square root of the sum of the squared differences of each dictionary element between pairs of papers[425]. Thus, for each pair of papers, one has a single number which is their proximity to each other, as defined by their dictionary scores. The greater this number, the greater the dissimilarity between papers. From this matrix of proximities[426] the PROXimity SCALing (PROXSCAL) algorithm was used to generate the co-ordinates for each paper in multidimensional spaces[427,428].

In the first instance, solutions were generated for 1-10 dimensions and the stress on the solutions calculated, in order to determine the most suitable number of dimensions for the analysis[429]. One can think of stress as providing a measure of how

---

[423] Note that the MDS takes the ~200 individual dictionary entries as variables, not the dictionary categories. The results of the MDS are therefore independent of the categorisation of the dictionary entries.

[424] The use of alternative metrics, e.g. city block, was not investigated.

[425] That is:

$$d(a,b) = \sqrt{\sum_{i=1}^{n}(a_i - b_i)^2}$$

[426] In fact, in *PASW 18* Proxscal calculates the proximities itself from the dictionary data.

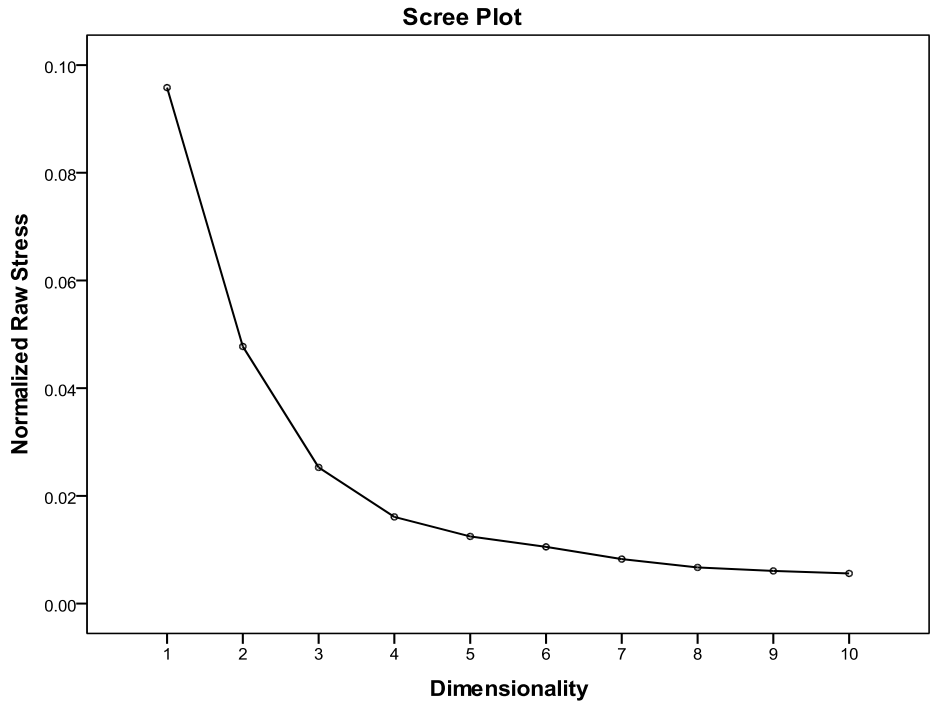
[427] Proxscal generates a least-squares representation of the proximity data. Initial conditions for the Proxscal were as follows: analysis was ordinal and used a Euclidean metric. The initial configuration began from a simplex start (because 'single random start' produced a local minimum), and the solutions were rotated (varimax). For further details about the Proxscal algorithm, see: (Borg & Groenen, 1997) pp. 432-433 see also: <http://www.scribd.com/doc/6886643/proxscal1>

[428] To be clear on the difference between proximities and distances: the proximities are calculated from the dictionary scores, and then the proxscal algorithm arranges these in a configuration in a multidimensional space. The distances between points in the configuration are not directly proportional to the proximities between papers.

[429] See: (Kruskal & Wish, 1978) pp. 53-56

much one must 'force' the proximity data into a configuration of a given dimensionality: high stress demonstrates disparity between the proximity data and the distances between points in the configuration. One aims to balance stress against solutions with fewer dimensions. The stress on each solution is shown on the scree plot in figure 6.10:

FIGURE 6.10: Scree Plot showing stress on 1-10 dimension solutions



Beyond three dimensions the decrease in stress gained from the marginal dimension is small, and the stress on the three dimensional solution is itself low. Additionally a three dimensional space is easier to visualise than higher dimension spaces. As a result the corpus was plotted in three dimensions[430].

Two dimensional projections of the configuration can be seen in the following figures (6.11-6.13), some outer points have been labelled to aid visualisation (table 6.14). As an initial attempt to interpret the three dimensional configuration, the extent to which scores on each dimension correlate with scores in individual dictionary categories will now be considered. The full table of correlations with the dictionary categories is shown in table 6.15.

[430] For the full table of co-ordinates, see the Corpus Data Spreadsheet <http://goo.gl/IqAh7>

FIGURE 6.11 Dimension 1 against Dimension 2

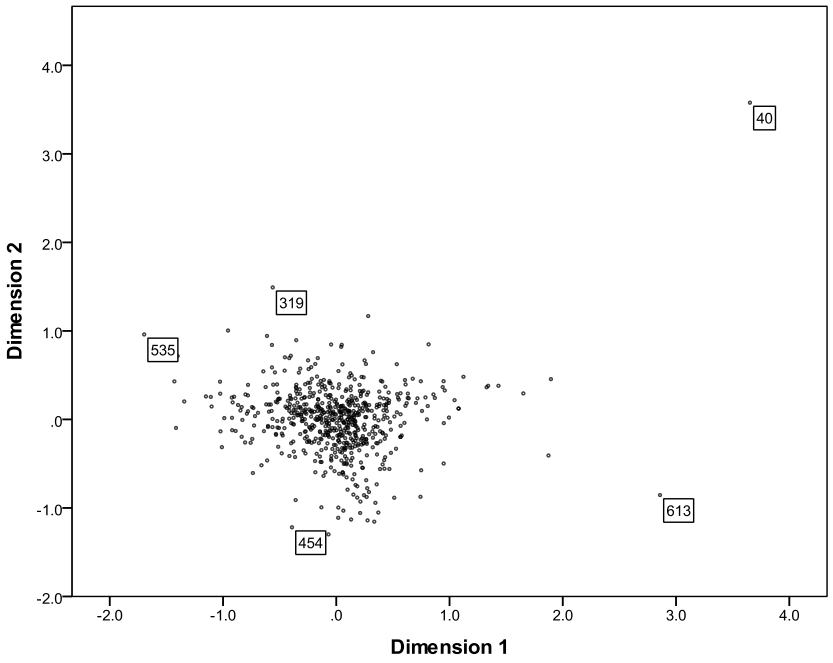


FIGURE 6.12 Dimension 1 against Dimension 3

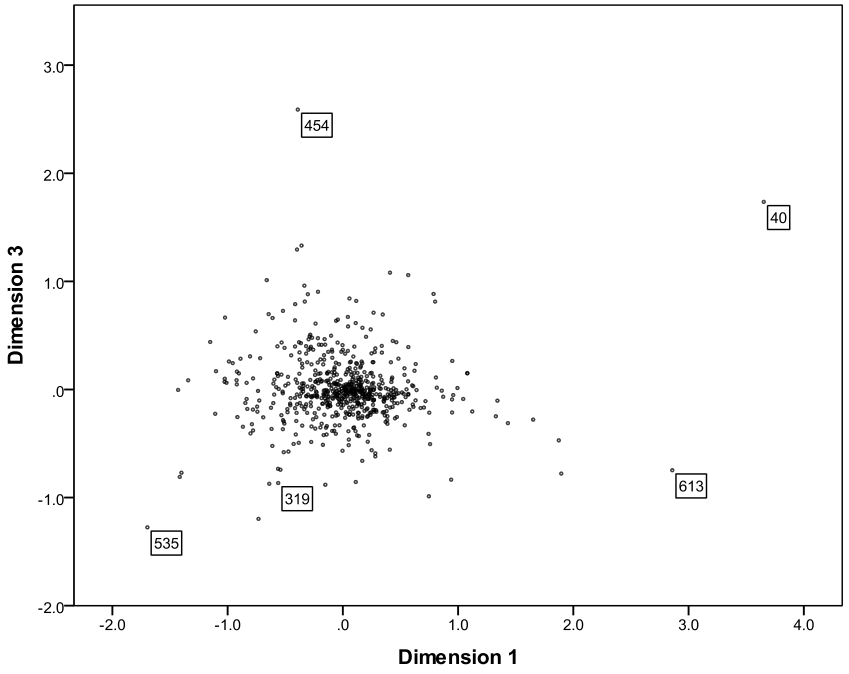


FIGURE 6.13 Dimension 2 against Dimension 3

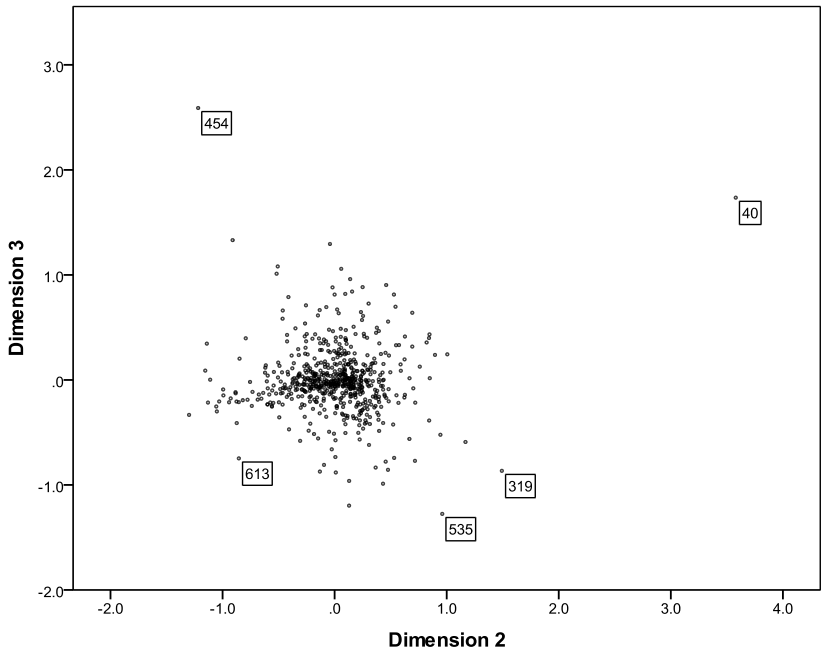


TABLE 6.14: Points labelled in Figures 6.11-13

Label No.	ID No.	Authors	Publication Date	Title	Journal
40	3301	A. B. S. Mitchell	1995	Evidence-based Medicine - accurate references are important	British Medical Journal
319	1796	O. Baenziger, H. U. Bucher	1999	Authority or evidence?	European Journal of Pediatrics
454	386	P. J. Graham, H. D. Dickinson	2007	Knowledge system theory in society: Charting the growth of knowledge-system models over a decade 1994-2003	Journal of the American Society for Information Science and Technology
535	4495	N. Cartwright	2010	What are randomised trials good for?	Philosophical Studies
613	51	R. Evans	2009	Evidence-based Orthopaedics or 'superstition in the pigeon'	Veterinary and Comparative Orthopaedics and Traumatology

TABLE 6.15: Correlations between dimensions and dictionary categories

		DIM	DIM_2	DIM_3
VIEWS	Pearson Correlation	.093*	-.253**	-0.03
TREATMENTS	Pearson Correlation	-.300**	.196**	-.084*
EFFECTIVENESS	Pearson Correlation	-.326**	.179**	-.146**
CONTEXT	Pearson Correlation	0.003	-.175**	.241**
PROFESSIONALS	Pearson Correlation	0.048	-.234**	.177**
PROBLEMATIC KINDS OF EVIDENCE	Pearson Correlation	-0.03	-0.07	0.014
PRINCIPLES	Pearson Correlation	.117**	-0.01	0.032
PHILOSOPHY	Pearson Correlation	.147**	-.354**	-0.04
PREFERENCES	Pearson Correlation	0.02	.103*	0.06
PATIENTS	Pearson Correlation	-.499**	.103*	-0.01
METHODS	Pearson Correlation	-.216**	.115**	-0.08
KINDS OF EXPERIMENT	Pearson Correlation	-.278**	.416**	-.123**
FAIR TEST CONCEPTS	Pearson Correlation	-.326**	.146**	-.143**
KNOWLEDGE, EXPERIENCE & SKILLS	Pearson Correlation	.166**	-.353**	.429**
KINDS OF EVIDENCE	Pearson Correlation	-0.02	.169**	.161**
IMPORTANT	Pearson Correlation	-.094*	0.08	.134**
EVIDENCE BASED MEDICINE	Pearson Correlation	.470**	-.612**	-.234**
DEALING WITH EVIDENCE	Pearson Correlation	.756**	.214**	-0.01
CRITICISM	Pearson Correlation	0.071	-.254**	-0.03
CAM	Pearson Correlation	-.170**	0.042	-.154**
AGAINST AND FOR	Pearson Correlation	.105**	-.310**	-0.08

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

6.2.3.1 Dimension 1

There is a strong correlation between dimension 1 (D1) and the "dealing with evidence" dictionary category (r=.756) (this correlation is stronger still for the single dictionary entry 'eviden*' within this category (r=.879)). There is also a moderate correlation with the "evidence based medicine" category (r=.470). As a consequence, D1 and these two categories share a similar pattern of negative correlations with other categories; as noted above in §6.2.2.2. Also, papers with a D1 co-ordinate greater than .7 all have the phrase EBM in their title, whereas the phrase occurs much less frequently in the titles of papers which score below -.7 on D1.

Standard Multiple Regression was performed, taking those dictionary categories for which $|r| > 0.3$ as the independent variables. As shown in table 6.16, below, these variables account for 74% of the variance on D1 (R Squared =.737). As we would expect "dealing with evidence" makes the largest unique contribution (beta=.594) and uniquely explains 26% of the variance in D1 (Part =.512). "Evidence based medicine" makes the next largest contribution (beta=.268). Other unique contributions were < 5%, but only the "treatments" category did not make a statistically significant unique contribution.

TABLE 6.16 Summary of Multiple Regression on D1

Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0.859*	0.737	0.735	0.243

Coefficients**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-0.55	0.055		-10.09	0	-0.66	-0.443					
	DEALING WITH EVIDENCE	3.931	0.159	0.594	24.692	0	3.619	4.244	0.756	0.706	0.512	0.74	1.348
	EVIDENCE BASED MEDICINE	2.175	0.194	0.268	11.201	0	1.793	2.556	0.47	0.412	0.232	0.75	1.331
	FAIR TEST CONCEPTS	-1.065	0.281	-0.09	-3.795	0	-1.62	-0.514	-0.326	-0.15	-0.079	0.77	1.305
	PATIENTS	-1.73	0.192	-0.213	-9.019	0	-2.11	-1.353	-0.499	-0.34	-0.187	0.77	1.302
	EFFECTIVENESS	-1.064	0.315	-0.082	-3.374	0.001	-1.68	-0.445	-0.326	-0.14	-0.07	0.73	1.37
	TREATMENTS	-0.001	0.247	0	-0.003	0.998	-0.49	0.484	-0.3	0	0	0.69	1.457

**Dependent Variable: DIM

6.2.3.2 Dimension 2

Dimension 2 (D2) correlates strongly and negatively with the "Evidence-based Medicine" category (r=-.612), and has further moderate negative correlations with the "philosophy" (r=-.354), "knowledge, experience & skills" (r=-.353), "against and for" (r=-.310) and "criticism" (r=-.254) categories. D2 also has a moderate positive correlation with the "kinds of evidence" category (r=.416). Within the "kinds of experiment" category, D2 correlates slightly better with the "randomised trials" subcategory (r=.419). D2 and "evidence-based medicine" share a similar pattern of correlations with other dictionary categories, but there are notable differences. For instance, D2 has a moderate correlation with "knowledge, experience & skills" (see above) whereas EBM is not correlated (r=.078[431]). Similarly "evidence based medicine" is negatively correlated with "patients" (see above), whereas D2 is only weakly correlated (r=.103)[432].

Also, inspection of the papers with high and low D2 co-ordinates (those which score |DIM_2| > .5) shows that, at the positive end, papers pick out discussions of randomised trials and the merits of particular experimental designs. Whereas at the negative end, papers talk about EBM in a much more general and reflective way.

Standard Multiple Regression was performed, taking those dictionary categories for which |r| > .25 as independent variables. As shown in Table 6.17 these variables account for 49% of the variance on D2 (R Squared=.491). "Evidence based medicine" makes the largest unique contribution (beta=-.529) accounting for 20% of the variance, followed by "knowledge, experience and skills" (beta =-.270).

[431] Not statistically significant at .05 level, two-tailed.

[432] Note that because "EBM" and dimension 2 are negatively correlated with each other, their respective correlations with "patients" are never the less in the same direction.

TABLE 6.17: Summary of Multiple Regression on D2

Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.701*	0.491	0.486	0.281

Coefficients**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	0.458	0.049		9.286	0	0.361	0.555					
	AGAINST AND FOR	0.45	0.77	0.02	0.584	0.559	-1.063	1.963	-0.31	0.024	0.017	0.688	1.454
	CRITICISM	-1.53	0.431	-0.112	-3.549	0	-2.376	-0.683	-0.254	-0.142	-0.102	0.834	1.198
	EVIDENCE BASED MEDICINE	-3.569	0.229	-0.529	-15.576	0	-4.019	-3.119	-0.612	-0.533	-0.449	0.721	1.388
	KNOWLEDGE, EXPIERENCE & SKILLS	-1.525	0.193	-0.27	-7.917	0	-1.903	-1.146	-0.353	-0.305	-0.228	0.714	1.4
	KINDS OF EXPERIMENT	0.172	0.164	0.039	1.048	0.295	-0.15	0.495	0.416	0.042	0.03	0.616	1.623
	PHILOSOPHY	-2.763	0.951	-0.095	-2.904	0.004	-4.631	-0.894	-0.354	-0.117	-0.084	0.774	1.292

** Dependent Variable: DIM_2

6.2.3.3 Dimension 3

Only one dictionary category is moderately correlated with D3. It has a positive correlation with “Knowledge, experience & skills” (r=.429). It also has a small positive correlation with the conceptually similar[433] “context” category (r=.241). Plus a small negative correlation with the “evidence base medicine” category (r=-.234).

Standard Multiple Regression was performed on these three dictionary categories (where |r| > .2) and together they explain only 27% of the variance in D3 (R Squared =.267). “knowledge, experience & skills” made the largest unique contribution (beta=.415), explaining 16% of the variance (Part =.394). The next largest contribution, from “evidence-based medicine” (beta=-.265) explained 7% of the variance (Part =-.264).

[433] That is to say, the two categories are moderately correlated with each other. I do not mean that 'conceptually similar' should be taken to imply that the relationship is necessary. It is contingent: it is a fact about how the EBM literature happens to emphasise and organise its concepts.

TABLE 6.18: Summary of Multiple Regression on D3

Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.517*	0.267	0.264	0.284

Coefficients**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-0.185	0.025		-7.346	0	-0.234	-0.136					
	EVIDENCE BASED MEDICINE	-1.514	0.198	-0.265	-7.66	0	-1.902	-1.126	-0.234	-0.295	-0.264	0.992	1.008
	KNOWLEDGE, EXPERIENCE & SKILLS	1.981	0.173	0.415	11.422	0	1.64	2.322	0.429	0.418	0.394	0.902	1.109
	CONTEXT	2.601	0.839	0.112	3.102	0.002	0.954	4.248	0.241	0.124	0.107	0.907	1.103

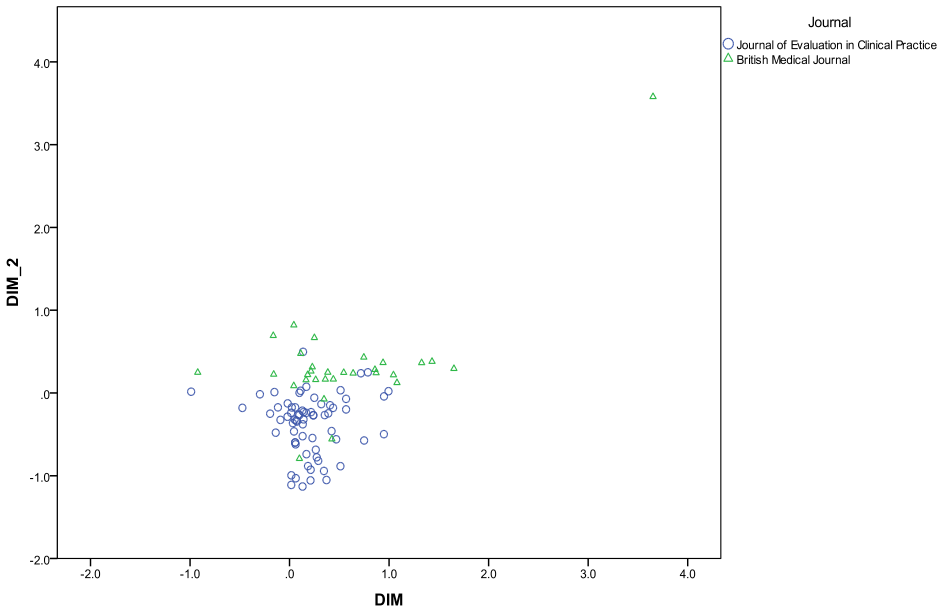
** Dependent Variable: DIM_3

6.2.3.4 Regions within the space

Some of the 11 most published in journals occupy particular regions in the three dimensional space.

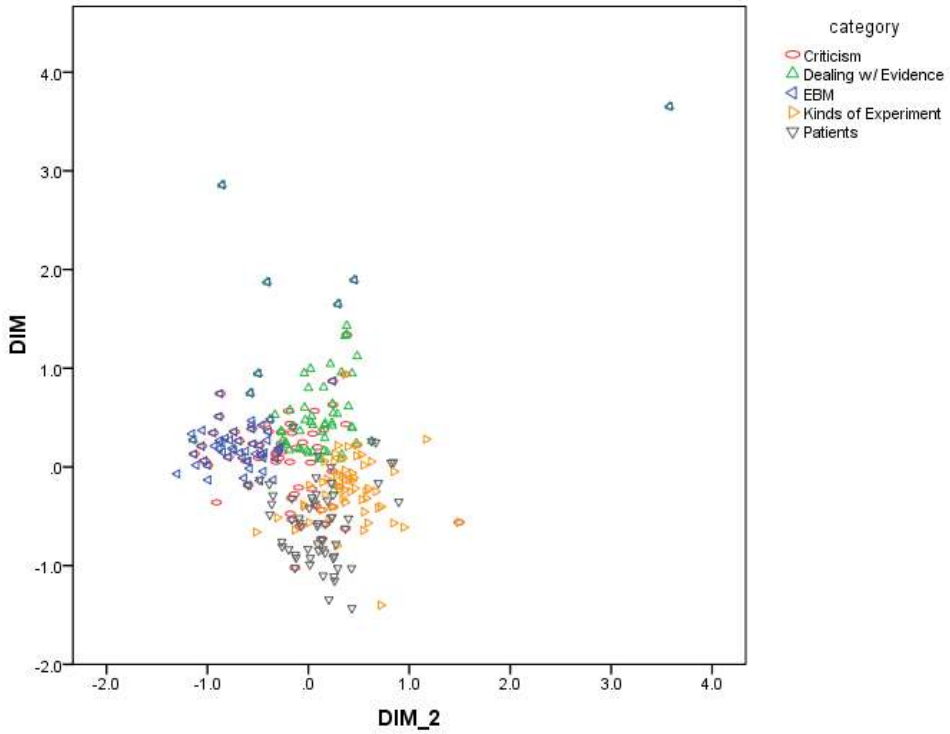
The BMJ and JECP both score positively on D1 (mean=.570 standard deviation=.788; m=.197 sd=0.31), but score quite differently on D2: the BMJ is positive on D2 (m=.342 sd=.684) whereas the JECP is negative (m=-.370 sd=.358). In contrast to the BMJ and JECP, the Journal of Clinical Epidemiology and the Annals of Internal Medicine both score low on D1 (m=-.481 sd=.509; m=-.501 sd=.406). Like the JECP, the journal Academic Medicine also scores negatively on D2 (m=-.266 sd=.209). With respect to D3, the Annals of Internal Medicine (m=.333 sd=.501), JAMA (m=.250 sd=.270), and Theoretical Medicine and Bioethics (m=.270 sd=.239) all score positively, whereas Perspectives in Biology and Medicine (m=-.149 sd=.149) scores negatively. Notably the Lancet occupies a central position in the space (D1: m=.060 sd=.825; D2: m=.072 sd=.447; D3: m=-.057 sd=.380). The different regions occupied by the BMJ and JECP are shown in Figure 6.19

FIGURE 6.19: D1 against D2, showing only papers from the BMJ and JECP



In addition to plotting particular journals within the space, papers that scored particularly highly within certain dictionary categories can also be plotted. For each dictionary category the highest scoring ten percent of papers were examined. Most high scoring papers in each dictionary categories cluster around the centre, however high scoring papers in five categories in particular occupy distinct regions within the space. This is shown in Figure 6.20:

FIGURE 6.20: D2 against D1, showing the ten percent highest scoring papers from five dictionary categories



6.3 Discussion

6.3.1 Interpreting the dimensions

6.3.1.1 Dimension 1

D1 is the most straightforward of the dimensions to interpret: it indicates the extent to which papers are explicitly about evidence.

First, the strong correlation between D1 and the "dealing with evidence" dictionary category and moderate correlation with the "evidence-based medicine" category show that D1 indicates the extent to which papers in the corpus talk about evidence directly[434]. Moreover the different regions occupied by the top ten percent of papers within five key dictionary categories, shown in Figure 6.20 confirm this. Figure 6.20 shows that the "dealing with evidence" category occupies the positive half of D1, whereas the "patients" category occupies the negative half. The former is precisely what one would expect if D1 captured the extent to which papers were explicitly about evidence. Together they also indicate an interesting conceptual opposition between talking about evidence, and talking about patients (see below). Thirdly, the higher scores on D1 of papers from the BMJ and JECP (Figure 6.19) provide another source of support. Those two journals are the two most published in journals in the corpus and both are known to be key sites in the literature for debate about EBM. The fact that they both score well on D1 demonstrates that a positive D1 co-ordinate represents more explicit engagement in debates about evidence in medicine. The fact that the Journal of Clinical Epidemiology also scores lower on D1 than other journals also fits well with this interpretation of D1, since clinical epidemiology clearly but *indirectly* occupies the same conceptual territory as EBM (see Chapter 5).

6.3.1.2 Dimension 2

The interpretation of D2 is less clear than it was for D1, however the most plausible interpretation is that it picks out the extent to which papers are critical and reflective about EBM (co-ordinate is negative), or emphasise clinical trials (positive). At the negative end of D2, papers talk about EBM in a general way and are more

[434] This is also confirmed by the fact that papers scoring highest on D1 all have the phrase EBM in their title; a fact which is not a necessary consequence of the analysis.

critical (as seen in the correlations with those categories, and in Figure 6.20). Papers at the positive end of D2 tend to be about randomised trials or trial design (also as suggested by the positive correlations noted above, and from Figure 6.20 showing the high scoring papers in the "kinds of experiment" category clearly within the positive region on D2). D2 indicates another interesting opposition: on D2, critical and reflective papers occupy a different space to papers about randomised trials.

This interpretation is also supported by the negative correlations of D2, with "evidence based medicine" and with the "philosophy", "criticism" and "against and for" categories; since these categories shift the discussion to a general level and introduce a more critical aspect. The good negative correlation of D2 with "evidence based medicine" is perhaps surprising since one might expect greater emphasis on EBM to go hand-in-hand with emphasis of randomised trials. On this interpretation however it is less surprising; the negative correlation suggests that the emphasis on EBM shifts the discussion to a more general level and indicates a more reflective stance.

Further support for this interpretation is given by the location of the BMJ and JECP again. On D2 the different positions of these journals are striking, as shown in Figure 6.19. Papers published in the BMJ are almost entirely confined to the positive end of D2, whereas papers published in the JECP are confined to the negative end. The JECP is known to hold a very critical stance towards EBM, hence its position at the more critical and reflective end of D2 confirms the interpretation. It would certainly be expected that the BMJ would hold a relatively orthodox view about EBM, and so its position too confirms the interpretation. Indeed, this also corroborates the provocative claim made by Beutow et al that: 'despite its long-held interest in EBM, the BMJ has never really interrogated the validity of this approach to clinical decision making[435]'.

6.3.1.3 Dimension 3

There are two plausible interpretations of D3. One interpretation of D3 is that it captures the extent to which papers are about the practical and experiential aspects of EBM. As with the interpretation of D2, D3 is better thought of as indicating

[435] (Buetow et al., 2006) p. 399

an opposition; in this case between EBM (at the negative end) and clinicians' experience (at the positive end).

This interpretation gets support from the positive correlation with the "knowledge, experience and skills" dictionary category, and negative correlation with the "evidence-based medicine" category. Some further support comes from the positioning of journals on D3, however D3 less clearly separates out particular journals. The Annals of Internal Medicine, JAMA, Social Science & Medicine, and Theoretical Medicine & Bioethics all score slightly higher on D3, if the interpretation is correct this indicates a greater emphasis on knowledge, experience and skills over EBM. This is less surprising in the cases of Social Science & Medicine and Theoretical Medicine & Bioethics, since these journals might be expected to offer a more contextualised perspective on EBM (although note that they place centrally with respect to D1 and D2). It is more surprising in the cases of Annals of Internal Medicine and JAMA, since these are mainstream general medical journals, and could be expected to occupy a similar space to the BMJ.

A second interpretation of D3 is that it fails to capture anything significant about the papers in the corpus. Note that the multiple regression on D3 yielded low R-Squared value, suggesting that the majority of the variation is noise (see table 6.18). It is arguable therefore that a two dimensional scaling solution would be equally appropriate for the data. Whilst the three dimensional solution does reduce stress (see figure 6.10) it is not clear that D3 captures any meaningfully interpretable aspect of the corpus. At a minimum, very little weight ought to be placed on the first interpretation of D3 suggested above, since the evidence for this is very weak.

6.3.2 What kind of support do the results give to the hypotheses?

Consider in turn the hypotheses stated at the beginning of this chapter:

(1) If the Categorical Interpretation is the dominant interpretation in the literature, then one would expect to find that discussion of EBM will be heavily focused on discussions of randomised trials.

There is no dictionary category corresponding only to randomised trials, however one might expect to see a strong relationship between discussion of EBM

and randomised trials to be apparent in the correlation between the "evidence-based medicine category" (and perhaps, to a lesser extent the "dealing with evidence" category) and the "kinds of experiment" or "fair test concepts" categories. One would expect moderate positive correlations. In fact, however, one sees small but significant negative correlations (see table 6.9).

The results of the MDS confirm this. High scores on D1 were interpreted as being about explicitness of engagement with evidence, and it is notable that this contrasted with low scores which showed emphasis on patients (Figure 6.20). D2 was interpreted as indicating an opposition between criticism of and reflection about EBM on the one hand, and emphasis on randomised trials on the other. While discussion of EBM is not heavily focused on randomised trials overall, it does seem that when discussion is less critical there is more emphasis on randomised trials. Similarly D3 was cautiously interpreted as indicating an opposition between clinical research and clinician's experience, which certainly is an opposition present in the Categorical Interpretation: these are the top and bottom, respectively, of the evidence hierarchy.

It may be argued that this does provide weak evidence for the dominance of the Categorical Interpretation in the literature. At least to the extent that the MDS demonstrates that the literature shares a similar conceptual structure: namely that orthodoxy is characterised by an emphasis on randomised trials and exclusion of patient's values as well as, to a weaker extent, that clinical research is to be contrasted with clinicians' experience. It is notable however that the evidence here is weak. I suggest that the results of the MDS are perhaps best interpreted as suggesting that the literature is conceptually messy and does not possess strong organising principles.

(2) If the literature contains many misperceptions and misrepresentations of EBM, then one would expect it to be very 'noisy', so there will be:

(2a) many different subsets of papers in the EBM literature, giving different interpretations of EBM.

(2b) one subset of the literature (perhaps in the top medical journals, or by prominent advocates of EBM) that represents the 'true' account of EBM.

The MDS shows no distinct clusters of papers. The corpus is arranged into a single, diffuse, central cluster (see figures 6.11-13). This is consistent with the hypothesis that the literature is 'noisy'. If we consider the variance on each dimension, for each year, there is no trend of increasing (or decreasing) variance. That is to say, the corpus is a single cluster in the space, which has shown similar levels of diffuseness over time. I claim that this suggests that the literature is *fundamentally unclear* about EBM, rather than simply full of misrepresentations. If EBM had been misrepresented, one would expect diffuseness to increase over time. Instead, it seems that the EBM literature has been noisy from the beginning. In general, we do not have good reasons to believe that the way EBM has always been talked about is any different from the way it is currently talked about. It would seem then that the EBM philosophy of evidence has remained as clear and as sophisticated as it ever was. Following the conclusions of Chapter 5, I claim it was never especially clear or sophisticated (see also hypothesis (3) below).

The interpretation of the dimensions given above does however suggest that certain regions of the space can be characterised, in broad terms. Firstly, the region that is positive on D1 and D2, and negative on D3 is where one would expect to find more orthodox papers about EBM; since one would expect papers in this region to be explicitly about evidence (due to their positive placement on D1), to emphasise randomised trials, rather than be critical (due to positive placement on D2), and to emphasise EBM over clinicians' knowledge, experience and skills (due to their negative placement on D3). Notably, the most highly cited EBM paper[436] 'Evidence-Based Medicine: what it is and what it isn't' falls within this region (Coordinates: .544, .247, -.071)[437]. Furthermore, almost all the papers in the corpus published in the BMJ fall within this region.

Secondly, the region that is positive on D1, and negative on D2 and D3 is where one would expect to find the papers that are critical of EBM; since one would expect papers in this region to be, again, explicitly about evidence (positive on D1), but more critical and reflective (negative on D2). One notable paper (and one of the 'key papers' used to generate the corpus) which occupies this region is 'The rise and

[436] (D. L. Sackett et al., 1996)

[437] Other key papers which occupy this region are: (Evidence Based Medicine Working Group, 1992; Rosenberg & Donald, 1995; Straus & McAlister, 2000)

fall of EBM[438]', written by prominent critics of EBM Bruce Charlton and Andrew Miles (Coordinates: .268, -.876, -.206)[439]. Additionally, almost all of the papers in the corpus published in the JECP score positively on D1 and negatively on D2, with a majority also then scoring negatively on D3.

(3) Given that EBM has been criticised and, so it has been claimed, evolved over the past twenty years, one would also expect to find temporal trends in the way that key concepts have been emphasised.

Neither the individual dictionary categories, nor dimensions 1 and 3 show significant correlations with the date of publication. D2 shows a small negative correlation with the date of publication (r=-.186), however the correlation is too small to be meaningful. Overall the MDS confirms that there is very little temporal structure to the corpus. This is surprising since many authors have claimed that EBM has increasingly acknowledged the role that patients' preferences and circumstances should play[440]. If we consider the corpus as a whole then there is no evidence that the "preferences" category scores have changed significantly over time. If however we consider the subset of papers by the most published proponents of EBM, then we find some evidence of such a trend: there is a moderate correlation between publication date and the "preferences" category (r=.524). Note additionally, that the subset of papers by the most published proponents of EBM also showed a moderate correlation between the date of publication and the "philosophy" category (r=.475). Otherwise, this subgroup is similar to the corpus as a whole. While it seems that some key proponents of EBM have put more emphasis on patients' values and have taken on a more philosophical orientation over time, this is not generally true of the corpus, or I claim therefore, the EBM literature.

[438] (Charlton & Miles, 1998)
[439] Other key papers which occupy this region are: (A. M. Cohen et al., 2004; Djulbegovic, G. H. Guyatt, & Ashcroft, 2009; Gupta, 2003; R Brian Haynes, 2002; Montori & G. H. Guyatt, 2008; Norman, 2001)
[440] For example: (Montori & G. H. Guyatt, 2008)(Howick, 2011)

6.4 Summary

I claim that the corpus, and therefore the EBM literature, presents a confusing picture of what EBM amounts to. Chapter 5 noted that the basic arguments for EBM were used to support very weak conclusions. The multidimensional scaling of the corpus highlights the room that these weak conclusions leave for further discussion of EBM. D2 and D3 capture interesting conceptual differences; namely the fact that discussion of evidence is opposed to talk about patients, that reflective discussion of EBM is opposed to talk about randomised trials, and to a lesser extent that discussion of evidence and EBM is opposed to more subjective talk about knowledge, experience and skills. If we add to this picture the fact that the corpus contains no temporal trends, then the EBM literature looks increasingly confusing. There is, I claim, no clear 'EBM view' reflected in the literature. This reinforces the need for critical clarification of what the EBM view should be. Nearly twenty years of literature has been surprisingly unhelpful in answering this question.

An alternative explanation of these results is suggested in the literature reviews that were discussed earlier in Chapter 5. The idea here was that the EBM literature as a whole is very noisy, because it is permeated by misunderstanding and misrepresentation. Never the less, it was claimed that behind the noise there is a coherent view which has been articulated, evolved and defended in the literature.

This explanation is harder to reconcile with the results presented here – I claim the EBM literature looks confusing because the concept is unclear; Straus et al[441] for example claim the EBM literature looks confusing because many authors are mistaken about EBM. If we wish to distinguish between the two explanations, then the key issue concerns whether some set of papers – without 'cherry picking' – can be put forward that plausibly characterise the EBM view.

The set of papers by the most published proponents of EBM in the corpus might be thought to fulfil this role, as might the set of key papers that was used to generate the corpus. Both are plausible candidates for where we might find 'the EBM view'. In both cases however, we see that these sets of papers paint broadly the same picture as the corpus as a whole. The only relevant differences being that the set of papers by the most published proponents of EBM shows a moderate

[441] (Straus & McAlister, 2000; Straus et al., 2007)

correlation between publication date and the “preferences” dictionary category. Neither of these two subsets represents a radically different position on EBM than we see in the corpus as a whole. In the absence of any other plausible subset of papers that might be thought to capture the EBM view, I conclude that we should instead think of the EBM literature as being simply unclear, rather than as hiding essence in noise. Whereas other authors have suggested that there is widespread misunderstanding of EBM, I claim there is flexibility of interpretation.

CHAPTER 7

7. How should evidence-based medicine be interpreted?

Uncertainty in the basic arguments for (Chapter 5) and literature about (Chapter 6) EBM means there is need for further specification of the EBM philosophy of evidence. As already noted, de-emphasising certain kinds of evidence, and stressing others, based on the vague concern that evidence 'may mislead' is not helpful. In so far as more detail is given, an epistemic reading of evidence hierarchies has already been described.

The results of Chapters 5 and 6 help to explain the difficulty of attaching a particular interpretation of EBM to its proponents, and the difficulty of assessing the dominance of those interpretations. The results from Chapter 5 revealed that, contrary to what some have claimed, there has been little change in how EBM has been talked about over the past twenty years. If there is room for ambiguity about whether the Categorical Interpretation truly applies, then there has always been room.

Unfortunately this creates a situation that further adds to the general confusion, since the details of EBM, and hence the kind of foundation it does, or does not, offer to the Canonical Criticism, seems even less clear. Even if it is unclear whether or not the Categorical Interpretation is representative of how EBM appears in the literature, there are still questions to ask about the role evidence hierarchies might play in filling in some of the epistemological details of the EBM view. The purpose of this chapter is to argue that the Categorical Interpretation is not a defensible interpretation of EBM. On a better interpretation, suggested by John Worrall and Jeremy Howick, it is argued that evidence hierarchies should not be read as revealing the epistemological details of EBM. In §7.1 I describe three problems with the Categorical Interpretation, in §7.2 I describe two solutions that have been proposed and in §7.3 I explore further one of those solutions, offered (independently) by John Worrall and Jeremy Howick.

7.1 Is the Categorical Interpretation defensible?

The Categorical Interpretation is not defensible[442]. However it is worth distinguishing two claims which the Categorical Interpretation might be thought to entail:

(1) Evidence from a given tier in the (relevant) hierarchy always provides more support (for the hypothesis in question) than evidence from lower tiers.

(2) Evidence from the top tiers of the (relevant) hierarchy always and only provides good support (for the hypothesis in question).

There are three well known problems with both (1) & (2).

7.1.1 The Bad Implementation problem

The first problem with both (1) and (2) concerns the way that studies employing a particular research design are implemented. (1) and (2) assume that the evidential weight assigned to a study is exhausted by fact that it employs a given research design. That is to say: particular randomised trials, in virtue of being randomised trials, are placed at the top of the hierarchy. Consider however that it is unclear what entities the tiers in the hierarchy refer to. If we think that the hierarchy ranges over actual studies, then (1) and (2) must be false for at least one obvious reason: there can be bad implementations of any particular research design. It is trivial to note that one could implement a highly ranked design badly: badly enough that it provides poor evidence (contrary to (2)), and badly enough that better evidence would be provided by a well implemented but lower ranked design (contrary to (1)). Grossman and Mackenzie[443] as well as Bluhm[444] have indentified this, seemingly trivial, point; noting that other authors have made exactly the mistake of ignoring it. Contrary to the Categorical Interpretation they point out that no one ought to hold the view that a badly implemented randomised study will always

[442] This has been argued in one form or another in, for example: (Bluhm, 2005; Borgerson, 2009; La Caze, 2008; Howick, Glasziou, & Aronson, 2009)
[443] (Grossman & Mackenzie, 2005)
[444] (Bluhm, 2010)

provide good support, and more support, for a hypothesis than an excellent observational study would.

Certainly if each tier in the hierarchy quantifies over actual studies, then the fact that a good observational study can provide better evidence than a bad randomised trial presents an uncontroversial counter-example to (1) and (2). One might argue that this is in fact a non-problem. In response therefore, one might argue that what is being ranked are *ideal* implementations. Actual studies need to be evaluated to determine whether they provide good evidence or not, but *in the ideal case* at least the hierarchy stands. That is to say, (1) and (2) should be understood as talking about properly implemented research designs; and it is these ideal cases that hierarchies quantify over in their rankings. Hence, the assumption of the Categorical Interpretation is merely that a properly-implemented randomised trial will always provide good evidence, and will also always provide better evidence than a well-implemented observational study.

7.1.2 The problem of Dramatic Effects

A second independent problem concerns the fact that good evidence for a hypothesis can, in fact, come from lower down the hierarchy. The problem is best illustrated in so-called 'dramatic effects' cases, such as in Smith and Pell's[445] famous paper about parachute use. Smith and Pell were (tongue in cheek) concerned about the hypothesis that parachutes are effective for preventing death and major trauma after freefall. They rightly point out that the evidence for this hypothesis is based, at best, on experience (not even any kind of *comparative* observations). They also, rightly, point out that this evidence constitutes very good evidence for the hypothesis: we know that parachutes are effective. Similarly other authors have, more seriously, presented examples where large effect sizes allow lower-tier evidence to provide strong evidential support for hypotheses[446]. Examples these authors give, where there is no doubt that a treatment caused a particular effect despite the fact that there is no comparative research supporting that causal claim, include: the 'Mother's Kiss' technique for removing blockages from a child's nose,

[445] (G. C. S. Smith & Pell, 2003)
[446] (Aronson & Hauben, 2006; Glasziou, Chalmers, M. Rawlins, & McCulloch, 2007) See also: (Howick et al., 2009)

laser treatment of Portwine stains, Fundoplication for heartburn[447], and oral ulceration resulting from the use of topical aspirin[448]. Recognising the problem of Dramatic Effects, Howick calls this the 'paradox of effectiveness'. He states: 'what we take to be our most effective therapies, ranging for the Heimlich manoeuvre to unblock an airway to eating to reverse the effects of starvation, have never been tested in randomised trials… it seems to follow that [on the Categorical Interpretation] our most effective therapies are not supported by "best" (randomised) evidence[449]'. Contrary to the Categorical Interpretation, these dramatic effects cases seem to show that one may have good evidence without having top-tier evidence.

Such examples clearly speak against (2), since they show that the threshold for good evidence can, in some cases, be set low down on the hierarchy. It is interesting to consider the way in which dramatic effects cases speak against (1). The claim made by (1) is about the relative strength of evidence; that evidence from higher up the hierarchy offers more support for a given hypothesis. In the parachute case this amounts to the claim that, although there is very good evidence that parachutes are effective at preventing death and severe trauma after freefall, *there would be better evidence* were there evidence from some comparative research; and better yet, a randomised study. Assessing this claim is complicated by the fact that the parachute example, and possibly the other examples, are arguably special cases within the subset of dramatic effects cases (because the probability of each hypotheses, conditional on one's 'total evidence'[450], is surely 1 – one knows[451]). Extra evidence doesn't help here (just as one does not need a tape measure to confirm (what one can see by looking) that an approximately 600 inch tree is not 6000 inches tall[452]).

[447] All three are from (Glasziou et al., 2007)
[448] From (Aronson & Hauben, 2006), who also present a series of other examples where anecdotal evidence of a particular effect can be, relatively unproblematically, attributed to a treatment (The examples all focus on adverse drug reactions).
[449] (Howick, 2011) p. 39
[450] Famously: 'In the application of inductive logic to a given situation, the total evidence available must be taken into account as a basis for determining the degree of confirmation' (Carnap, 1950) p. 211 See also: (Hempel, 1968) p. 125 and (T. Williamson, 2000) pp. 189-90
[451] And we are fallible in so far as we can be wrong about what our total evidence is. Indeed, Timothy Williamson argues that we are not always in a position to know what our total evidence is. See: (T. Williamson, 2000) and (T. Williamson, 2007) ch. 5
[452] This example is based on (T. Williamson, 1992)

The problem of Dramatic Effects perhaps does not count against (1). Outside of these special cases, where one *knows*, it seems plausible that evidence from higher in the hierarchy will always offer incrementally more support for a hypothesis. Of course if one has good evidence, that is, if one's evidence is sufficient for all practical purposes, then one has no need for 'better' evidence. As with any hypothesis however, it would always be better if one knew it.

7.1.3 The problem of Small Effects

A third problem, which speaks against both (1) and (2), concerns the fact that when effect sizes are small the probability of false positive results increases, no matter what research design is employed. The most dramatic illustration of this point is provided in another well-known paper, by Leonard Leibovici[453]. Leibovici performed a randomised trial investigating the effect of remote, retroactive, intercessory prayer on patients who had suffered from bloodstream infections. 3393 patients, treated for bloodstream infections between 1990 and 1996, were randomised to two groups in 2000, one of which was randomly chosen to be prayed for. Leibovici found that while intercessory prayer had no significant effect on mortality, the intervention group had a statistically significant shorter duration of fever, and a shorter stay in hospital (of course, these two outcomes are not entirely independent).

Putting aside any methodological comments on the study[454] the key point to note is that even high-quality studies will occasionally deliver false positives[455], especially if the actual effect size is small. This example is useful because that fact becomes clearly apparent. The hypothesis that the study provides a false-positive result is considerably more probable than the hypothesis that remote retroactive intercessory prayer is effective (because one's total evidence rules out the

[453] (Leibovici, 2001)

[454] For example, what was the prior justification for measuring those particular outcomes? - Given the nature of the result, one might suspect they were chosen post hoc precisely because they were significantly different; or, if there are good reasons for measuring these outcomes, that the randomisation was repeated until a statistically significant result was delivered. Although regarding the latter point, Leibovici states in the online comments to the paper that the randomisation was performed only once: <http://www.bmj.com/content/323/7327/1450.abstract/reply#bmj_el_20476>

[455] In other actual cases the fact that a particular result might be a false-positive is easy to overlook. In relation to this see, notably: (Ioannidis, 2005)

effectiveness of retroactive prayer). The example is such that one knows it must be a false-positive.

Leibovici's study illustrates two points which are relevant to this discussion. First, contrary to (2), evidence from higher tiers does not always provide good evidence. Despite being a (putatively) well-implemented randomised trial, it does not constitute good evidence for the hypothesis. Moreover, on the assumption that it is free from methodological problems, the primary reason to reject Leibovici's result is based on mechanistic evidence. It is, presumably, the knowledge that one cannot cause events that happened in the past, and the knowledge that there is no mechanism by which remote intercessory prayer can cause therapeutic effects, which (as parts of one's total evidence) justify the belief that Leibovici's result is a false-positive. Thus, the second point to note is that this reasoning involves using mechanistic evidence to defeat clinical research evidence. As noted above, mechanistic evidence is univocally placed at the bottom of hierarchies, if it is placed at all. Hence, contrary to (1), evidence for lower down the hierarchy in this case defeats evidence from higher up.

The two effect size problems show that a Categorical Interpretation of evidence hierarchies, in either sense (1) or (2), cannot be sustained. Contrary to the Categorical Interpretation, evidence at the top can be poor, evidence lower down can be better than evidence higher, and evidence at the bottom can be good. In the next section I describe two attempts to solve these problems.

7.2 Solutions to the problems

The Categorical Interpretation of evidence hierarchies possesses the magic combination of being both explicit and naïve. In the philosophy of science literature, there have been two explicit responses to the Categorical Interpretation and its problems; the first, owing to Adam La Caze[456], and a second owing to John Worrall[457] and Jeremy Howick[458]. These will be discussed in turn:

La Caze argues primarily for greater limits on the scope of evidence hierarchies, most importantly he argues that evidence hierarchies should not been seen as ranking research designs according to the level of evidential support they

[456] (La Caze, 2008, 2009) See also: (La Caze, 2011; La Caze et al., 2011)
[457] (Worrall, 2002, 2007b)
[458] (Howick, 2011; Howick et al., 2009)

give to medical claims or clinical decisions[459]. His re-interpretation of the hierarchy would seem to be no less categorical however; he states:

> 'EBM's hierarchy should be interpreted as a hierarchy of comparative internal validity… all other things being equal, studies that utilize the methods higher in EBM's hierarchy have higher internal validity than studies designed according to the methods lower down the hierarchy[460]'.

La Caze preserves a categorical ranking, but changes the account of what is being ranked (other things equal, higher ranked designs possess greater internal validity). Research designs are not ranked according to the evidential support they lend to hypotheses, but rather, according to the relative level of internal validity they possess. This re-interpretation solves both the bad implementation and the effect size problems by pulling apart the link between position on a hierarchy and levels of evidential support. As La Caze argues, the task of showing that a given study supports (or not) some hypothesis requires further argument on top of an evaluation of its internal validity. The fact that one can rank study designs according to their relative levels of internal validity does not, on its own, entail anything about the evidential support that those research designs may lend to a hypothesis. In general then, La Caze argues that a higher level of internal validity in a study is not synonymous with that study providing greater evidential support.

Worrall and Howick offer a rather different solution. Howick argues that the categorical claims, (1) and (2), should be replaced with *empirical* claims: as a matter of fact, top level evidence is *often or generally* the best evidence, lower tier evidence is often poor evidence, and evidence from higher tiers is often better than lower tier evidence[461].

This is in danger of repeating the problems with the basic arguments for EBM, given above, namely of being trivial and vague. Randomised trials may often provide the best evidence, the question is precisely *when*? Worrall and Howick can

[459] (La Caze, 2008) See also: (La Caze, 2011)
[460] (La Caze, 2009) pp. 2-3
[461] (Howick, 2011) p. 4

be considered together because they both offer essentially the same response to this further question. Worrall claims:

> 'Best evidence for the positive effect of a therapeutic intervention arises when plausible alternative explanations [of that effect]… have been eliminated[462]'

And Howick claims:

> '[we should replace] the categorical ranking of randomised trials above observational studies with the requirement that in order to accept that a treatment has clinically relevant effects, the treatment must demonstrate an effect that outweighs the combined effect of plausible confounders[463]'

Both Howick and Worrall are making a point about discriminating between alternative hypotheses. This idea will be discussed further below, but it is important to note first that La Caze's re-interpretation of the hierarchy and Howick's proposal for replacing the hierarchy are independent of each other: the two views are compatible. Both argue against interpreting evidence hierarchies categorically. Whereas Worrall and Howick put forward a view about what counts as good evidence that is independent of one's interpretation of evidence hierarchies, La Caze puts forward an interpretation of evidence hierarchies that is independent of one's view about what counts as good evidence. Together they present two complementary arguments for the same conclusion: evidence hierarchies should not be interpreted categorically. More generally, they both argue that evidence hierarchies should not be given an epistemic reading. That is to say, evidence hierarchies should not be read as if they supplied a theory about what counts as good evidence for medical claims or clinical decisions.

[462] (Worrall, 2002) p. S328
[463] (Howick, 2011) p. 119, see also p. 40.

Importantly for this discussion, Worrall and Howick do offer something like an epistemological theory that deals with the three problems, above[464]. Since Worrall and Howick aim to give an account of what counts as good evidence, it is that idea that will be the focus of the following discussion. Below I describe how Worrall and Howick's view solves the three problems, above: firstly I discuss the solution to the effect size problems, then the bad implementation problem.

The better one can discriminate between and thereby rule out alternative hypotheses the better evidence one has for one's own hypothesis. Consider how this view put forward by Worrall and Howick's view escapes the two effect size problems.

The dramatic effects examples represent perhaps the ideal case for the view because in those examples the evidence is almost perfectly discriminating; the evidence (for example, very high survival rates from freefall with a parachute, very low without) can plausibly only be accounted for by one hypothesis (the effectiveness of parachutes). The important point about dramatic effects cases is not that the effect size is absolutely large, rather it is that one can detect such large and dramatic effect sizes even with methods that have a substantial margin for error[465]. Or put the other way round: even methods with a substantial margin for error can detect effects, if the effect size is large enough.

It is perhaps also worth noting that this view does not entail that when effect sizes are large one can get away with 'weaker' evidence. If by 'weaker' one means evidence from studies lower down the hierarchy. On Worrall and Howick's view, whether or not evidence is weak depends on whether the observed effect size is comparable with the effect that bias and error could have had; that is, whether there is some plausible alternative explanation the result. The point is that when one can 'demonstrate an effect that outweighs the combined effect of plausible confounders[466]', then one has good evidence; if one can truly demonstrate this with the method one used, then it is irrelevant if that method happens to be ranked low

[464] Most explicitly, the principle is just that: we have good evidence when the 'effect size outweighs the combined effect of plausible confounding factors' (Howick, 2011) p. 40

[465] Consider Timothy Williamson's tree example: by eye, one's estimates of a tree's height in inches are not accurate: the margin for error is substantial. However one's judgements, by eye, can still provide good evidence for *some* hypotheses about the tree's height. The difference between 600 inches and 6000 inches is sufficiently 'dramatic' that even by eye it can be discriminated. When looking at a 600 inch tree therefore, one has good evidence for the hypothesis that the tree is not 6000 inches; even though one's eyes are not a good method for judging heights of trees in inches. (T. Williamson, 1992)

[466] (Howick, 2011) p. 119

on evidence hierarchies, or if it is more prone to bias - one is *not* 'getting away with weaker evidence' in that case.

The small effects problem is also easily dealt with. In the case of small effects the risk of false-positives increases because – even with methods that have a smaller margin for error – bias and error can plausibly account for the small effect size. That is to say, the evidence fails to discriminate between one's evidence being the result of the effect of bias and error, and one's evidence genuinely being the result of the effect of the intervention. Again, the important point is not the absolutely small effect size, but the fact that the effect size is on the limit of a method's resolution. Even the most accurate methods are not perfectly discriminating.

Worrall and Howick's view is not controversial, at least at this level of explanation. Howick, for example, claims that the view is based on an uncontroversial 'scientific common sense' intuition[467]. As further illustration, it should be noted that other authors have also made use of the same idea. To take an example from another philosopher of science speaking specifically about medicine, Alexander Bird has recently argued that Austin Bradford Hill's 'criteria of causation[468]' can be unified and explained with reference to the notion of ruling out alternative hypotheses[469]. Explicitly he states: 'a good criterion of causation is one that, when fulfilled, succeeds in eliminating potential error, i.e. it eliminates an alternative, false hypothesis[470]'. Furthermore the insight can be expressed in other more formal terms, too. For example, Sherylin Roush puts forward an account of evidence explicitly based on the idea of evidence being discriminating[471]. On Roush's view, one compares the probability of seeing the evidence, given the hypothesis in question is true, with the probability of seeing the same evidence, if that hypothesis is false[472]. When there is much bias or error (or equally, when there are multiple plausible explanations of the evidence), the probability of seeing that same evidence, given the hypothesis is false is raised. This brings the likelihood ratio (P(e|h)/P(e|-h)) closer to one, indicating that

[467] (Howick, 2011) p. 33
[468] See: (Hill, 1965)
[469] (Bird, 2011) see relatedly: (Bird, 2005). Note also that similar points have also been made elsewhere, for example: (Howick et al., 2009)
[470] (Bird, 2011) p. 242
[471] In fact, she claims that good evidence is a 'discriminating indicator' of a hypothesis, however it is only her 'discrimination condition' that is mentioned here, in passing. See: (Roush, 2005) Ch. 5.
[472] There are a number of different (non-trivially so) ways to formalise this idea (Eells & Fitelson, 2001). The likelihood ratio (or the ordinally equivalent log of it) has some desirable formal features, and is the formalisation defended in (Roush, 2005) Ch. 5.

the evidence is less discriminating[473]. The point of these two examples is simply to show that Worrall and Howick's idea is not a novel one to philosophers of science or epistemologists.

In addition to the two effect size problems, consider second that the Bad Implementation problem does not arise on Worrall and Howick's view. In a sense, the reason is trivial; their account of good evidence is supposed to replace the account of good evidence provided by the Categorical Interpretation. If there is no hierarchy in one's account of good evidence, then there is no need to worry about what it quantifies over. It is however interesting to note where the issue of implementation fits into their account. Rather than relying merely on research design as a guide to evidential support, Worrall and Howick's account entails that assessing the implementation of the research design is necessary for determining whether one has good evidence. The reason is that, to take one example, flaws in a study introduce alternative explanations of the results. For example, failure to randomly allocate patients to experimental groups introduces the possibility of selection bias. It is this feature of their view that will now be discussed.

On the basis of Worrall and Howick's view, I intend to argue for two claims: first, that their view re-locates the epistemological details of EBM in the techniques of critical appraisal. Second, that evidence hierarchies should be given a heuristic reading. These two claims will be of use in the evaluation of the tension in the arguments made by opponents of homeopathy, between the interpretation of EBM and the use of mechanistic evidence.

7.3 Re-interpreting evidence-based medicine

7.3.1 Critical appraisal

The processes of critically appraising an article from the literature is described in a number of different ways in EBM textbooks. In the *Users' Guide to the Medical Literature* (series[474] and book[475]) a three step process for using an article

[473] *That* the evidence is less discriminating when P(e|-h) is raised is vague enough to be true on any formalisation.

[474] A series of 32 papers published between 1993 and 2000 in *JAMA*: (Barratt et al., 1999; Bucher, G. H. Guyatt, D. J. Cook, Holbrook, & McAlister, 1999; A. L. Dans, L. F. Dans, G. H. Guyatt, & W. S. Richardson, 1998; Drummond, W. S. Richardson, O'Brien, M. Levine, & D. Heyland, 1997; GH et al., 1995; Giacomini & D. J. Cook, 2000a, 2000b; G. H. Guyatt & D

from the literature is proposed. Each step focuses on answering a separate question: first, 'are the results of the study valid?'; second, 'what are the results?'; and third 'how can I apply these results to patient care?'[476]. In the first edition of the EBM textbook *How to Practice and Teach EBM* critical appraisal of evidence is given as a two step process: 'deciding whether it is valid… and deciding whether it is important[477]'. In more recent editions of the same textbook the practice of EBM is broken down into five steps, step three of which entails: 'critically appraising… evidence for its validity… impact… and applicability[478]. These differences are not substantive however and the common element is clear: critical appraisal is a set of techniques for evaluating whether results from a study constitute good evidence for some medical claim and whether that claim is useful for making a particular clinical decision. In the textbooks and guides critical appraisal is often presented as a set of salient questions one should ask of a given study.

In Chapter 5 it was argued that critical appraisal cuts across the distinction between EBM considered as an account of evidence for medical claims or for clinical decisions. It rightly applies to both however for the present discussion critical appraisal can be narrowed. The application of evidence to clinical decisions will not be considered here; rather, the interest is in determining whether some evidence constitutes good evidence for a particular medical claim. I suggest that the account of good evidence put forward by Worrall and Howick describes perfectly the underlying epistemic purpose of the techniques of critical appraisal.

There is of course variety in the techniques one must apply in critically appraising evidence. Note that just as there are different hierarchies for claims of treatment benefit, harm, and diagnostic test accuracy etc, the set of specific

---

Rennie, 1993; G. H. Guyatt et al., 2000, 1997; G. H. Guyatt, Jack Sinclair, D. J. Cook, & Glasziou, 1999; G. Guyatt, D. Sackett, & D. Cook, 1993, 1994; Hayward, M. C. Wilson, Tunis, Bass, & G. H. Guyatt, 1995; Hunt, Jaeschke, & McKibbon, 2000; Jaeschke, G. H. Guyatt, & D. L. Sackett, 1994a, 1994b; A Laupacis, Wells, W. S. Richardson, & P Tugwell, 1994; M. Levine et al., 1994; McAlister, Andreas Laupacis, Wells, D. L. Sackett, & for the Evidence-Based Medicine Working, 1999; McAlister, Straus, G. H. Guyatt, & R Brian Haynes, 2000; McGinn et al., 2000; Naylor & G. H. Guyatt, 1996a, 1996b; Oxman, D. J. Cook, & G. H. Guyatt, 1994; O'Brien, D. Heyland, W. S. Richardson, M. Levine, & Drummond, 1997; Randolph, R Brian Haynes, Wyatt, D. J. Cook, & G. H. Guyatt, 1999; W. S. Richardson & Detsky, 1995a, 1995b; W. S. Richardson, M. C. Wilson, G. H. Guyatt, D. J. Cook, & Nishikawa, 1999; W. S. Richardson, M. C. Wilson, J. W. Williams, Moyer, & Naylor, 2000; M. C. Wilson, Hayward, Tunis, Bass, & G. H. Guyatt, 1995)

[475] (G. H. Guyatt & Drummond Rennie, 2002)

[476] (G. H. Guyatt & Drummond Rennie, 2002) p. 76-7

[477] (D. L. Sackett et al., 1997) p. 80

[478] (Straus et al., 2005) p. 4

questions one asks when appraising a study will differ according to the kind of claim the results are supposed to be evidence for[479]. Equally, questions will also differ depending on what kind of research design was employed. The appraisal of a randomised trial looking at treatment benefit requires one to ask different specific questions than the appraisal of an observational study looking at treatment benefit. Indeed textbooks on critical appraisal often divide their sections according to either different kinds of claim, or different kinds of research design that one might appraise[480]. More explicitly the following table illustrates different kinds of critical appraisal questions, one should ask[481]:

[479] (G. H. Guyatt & Drummond Rennie, 2002) pp. 33-37

[480] So for example, the contents pages of (G. H. Guyatt & Drummond Rennie, 2002), (Straus et al., 2005) and (Trisha Greenhalgh, 2006) show that the books are divided into sections according the kind of medical claim (diagnosis, prognosis, therapy, harm, economic evaluation etc). And in contrast (Crombie, 2008) is divided into sections according to the kind of research design employed.

[481] Adapted from tables 3.2 (p. 71), 5.1 (p. 117) and 6.1 (p. 178) in (Straus et al., 2005)

TABLE 7.1: Critical Appraisal questions for appraising randomised trials looking at treatment benefit, studies looking at treatment harm, and studies looking at the accuracy of diagnostic tests.

Treatment Benefit	Treatment Harm	Diagnostic tests
Was the assignment of patients to treatment randomised?	Were there clearly defined groups of patients, similar in all important ways other than exposure to the treatment or other cause?	Was reference to ("gold") standard measured independently?
Was the randomisation concealed?	Were the treatments/exposures and clinical outcomes measured in the same ways in both groups? (was the assessment of outcomes either objective or blinded by exposure?)	Was the diagnostic test evaluated in an appropriate spectrum of patients?
Were the groups similar at the start of the trial?	Was follow-up of patients sufficiently long and complete?	Was the reference standard ascertained regardless of the diagnostic test result?
Was follow-up of patients sufficiently long and complete?	Is it clear that the exposure preceded the onset of the outcome?	If one is concerned with a cluster of tests of clinical prediction rules, was the cluster of tests validated in a second, independent group of patients?
Were all patients analysed in the groups to which they were randomised?	Is there a dose response gradient?	
Were patients, clinicians and study personnel kept blind to the treatment?	Is there any positive evidence from a 'dechallenge-rechallenge' study?	
Were groups treated equally, apart from the experimental therapy?	Is the association consistent from study to study?	
	Does the association make biological sense?	

I claim that while the specific appraisal questions may differ, the *epistemic aim* is always the same. Indeed the epistemic aim just is that suggested by Worrall and Howick: it is to determine whether the evidence discriminates between plausible alternative hypotheses. The appraisal questions in table 7.1 highlight particular ways the result can be confounded. To pick two examples: First, in a study of treatment benefit one asks whether patients were randomised, because this rules out selection bias (by definition) and because it reduces the plausibility that threats to internal validity are confounded with the treatment[482]. Second, in any comparative research one asks whether groups were similar at the start of the trial because, if any relevant dissimilarities exist, this may introduce further confounding to consider.

A simple example from the *Users' Guide* provides further illustration of the idea that critical appraisal represents the operationalisation of Worrall and Howick's account of good evidence:

> 'Consider the question of whether, in very sick people, hospital care prolongs life. A study finds that more people die in hospital than in the community. We would easily reject the naïve conclusion that hospital care kills because, intuitively, we understand that hospitalised patients are generally much sicker than patients in the community[483],

Guyatt and Rennie read this example as showing that the evidence, 'more people die in hospital than in the community', would seem to support the hypothesis, 'hospital care kills'; until, that is, it is recognised that the evidence is insensitive to any difference in sickness levels in the two settings. The claim is that the method used – mere counting of deaths in the two settings – cannot discriminate between greater deaths arising from hospital care, and greater deaths arising from hospital patients being systematically more likely to die: since one would see the same evidence (higher counts, compared to the community) in either case.

It should be noted that as the effect size problems discussed above demonstrate, whether Guyatt and Rennie are correct here depends on just *how deadly* hospital care is. How many more people do die in hospital? - If hospital care

[482] (Shadish, T. D. Cook, & D. T. Campbell, 2002) pp. 248-251
[483] (G. H. Guyatt & Drummond Rennie, 2002) pp. 86-7

were truly deadly, such that even relatively healthy patients died, then one could likely detect this through mere counting; despite there being systematic differences in sickness levels between hospital and community settings. The empirical assumption underlying Guyatt and Rennie's simple example is that – as one knows – hospital care is not so deadly that it will be clearly discriminable from greater deaths merely due to systematic differences in sickness levels. Or put another way, their assumption is that hospital care is safe enough to be confusable with differences in sickness levels, if one merely counts the number of deaths between hospital and community settings.

This example again demonstrates that when one critically appraises a study – even as in this simple example – the aim is to determine whether the evidence discriminates between plausible alternative hypotheses. The aim is not simply to indentify whether there are confounders, but to evaluate whether those confounders are therefore likely to provide an alternative explanation of the evidence. As this example also demonstrates critical appraisal takes place against a background of judgements about what is a plausible alternative explanation and whether counting deaths is likely to provide sufficiently discriminating evidence.

Worrall and Howick's account of good evidence describes the epistemic aim behind techniques of critical appraisal.

7.3.2 The heuristic interpretation of evidence hierarchies

The argument above suggests two further questions: what justifies the claim that evidence from randomised trials is *often* the best evidence? How should evidence hierarchies be read, if they are not the source of EBM's epistemological details? The answer to the first gives the clue to the second.

On Worrall and Howick's view, the justification for the claim that randomised trials often provide the best evidence for treatment effects must rest on the empirical claim that, as a matter of fact, the magnitude of the effect sizes of most treatments are often on a par or smaller than the magnitude of the biases inherent in other methods. The greater discriminatory power of randomised trials is necessary only if that is the case. It is the greater discriminatory power of randomised trials, coupled with the contingent fact that such levels of discrimination are most often what are needed, that means randomised trials often provide the best evidence.

Indeed, Worrall quotes a letter, published in 1980, in the BMJ from Richard Doll and Richard Peto where they provide precisely this justification:

> 'therapeutic advances over the past decade or so have involved recognition that some particular treatment for some condition yields a moderate but important improvement in the proportion of favourable outcomes[484]'

Glasziou et al, for example, have also reiterated this point, more recently:

> 'randomised trials will remain the principal means of obtaining reliable evidence about the average effects of treatments when effects are moderate[485]'

In answer to the first question, what justifies the claim that randomised trials often provide the best evidence, Worrall and Howick's answer is that the greater discriminatory powers of randomised trials are needed because the effect sizes of treatments are often moderate. Consequently, this suggests an answer to the second question, how should evidence hierarchies be read, if they are not the source of EBM's epistemological details. Rather than reading anything epistemologically significant into hierarchies of evidence, I claim that Worrall and Howick's view of evidence implies that evidence hierarchies are *heuristically useful*.

Consider that in evaluating some medical claim, one ought to critically appraise the total evidence for that claim. If however one knows that randomised trials often provide the best evidence for the claim that a treatment has some putative effect, because such effects are mostly only moderate effects, then one has reason to narrow one's evidence base by focusing only on the appraisal of randomised trials. If one is also constrained by practical factors, such as time, then such reasons come into play.

Hierarchies of evidence, if they are read heuristically, provide an epistemologically crude, but practically useful way to pare down one's total evidence into a more manageable body of evidence to appraise, whilst also minimising the

[484] (Doll & Peto, 1980) See: (Worrall, 2002) p. S325
[485] (Glasziou et al., 2007) p. 351

epistemic compromise. Note that this heuristic reading of evidence hierarchies highlights their scope and limits. Firstly, considering the limits, it shows that there is nothing epistemologically deep contained in hierarchies. They are like approximations; just as the small angles approximation does not reveal anything mathematically deep about trigonometric functions[486]. Secondly, considering the scope, a heuristic reading of evidence hierarchies emphasises their contingency on particular, resource constrained, circumstances.

The move from an epistemic to a heuristic reading of evidence hierarchies has important consequences – for one, it provides a better way to interpret some of the 'categorical' statements in the EBM literature. Consider again a favourite quote of those who argue that the EBM literature supports a Categorical Interpretation:

> 'If the study wasn't randomised, we suggest that you stop reading it and go on to the next article in your search (Note: We can begin to rapidly critically appraise articles by scanning the abstract to determine if the study is randomised; if it isn't we can bin it). Only if you can't find any randomised trials should you go back to it.[487]'

Clearly *something* epistemologically interesting is expressed here: randomisation is considered a methodological virtue. It is randomised trials that one is instructed to look at first: one is permitted to 'bin' the rest, only coming back to them in the absence of randomised trials.

Consider however why it is that Straus et al suggest one should stop reading non-randomised studies. On an epistemic reading of evidence hierarchies, the answer is that non-randomised studies can be 'binned' because the evidence they provide is categorically worse. Worrall seems to think this quotation does express such a view, even if he goes on to claim that proponents of EBM no longer endorse that view[488]: I claim he is mistaken.

On a heuristic reading of evidence hierarchies, this quotation can be given a more reasonable interpretation. The heuristic interpretation easily accommodates the idea that Straus et al's advice is premised on the fact that one only has a limited

[486] To push the analogy further (too far?): one might say that critical appraisal of the total evidence is like a Taylor series expansion.
[487] (Straus et al., 2005) p. 118
[488] (Worrall, 2007b)

time to critically appraise evidence (indeed one must does so 'rapidly'): the constraint that motivates the advice is not primarily epistemological, it is practical. Given the practical constraints a clinician faces when finding and appraising evidence a crude hierarchy-heuristic is an invaluable tool.

In fact another EBM textbook previously mentioned, the *Users' Guide*, also contains numerous statements indicating that practical factors play the biggest role in constraining clinician's selection of evidence when they critically appraise the literature; for example:

> 'The biggest challenge to evidence-based practice: time limitation[489]'

> '[clinicians] often feel overwhelmed by the magnitude of the medical literature. Evidence based medicine offers some solutions to this problem[490]'

> 'Because our time for searching is limited, we would like to ensure that there is a good chance our search will be productive[491]'

Thirdly and more recently, the Oxford Centre for Evidence Based Medicine in 2011 published a version of an evidence hierarchy - the 'levels of evidence #2' table[492] – where the heuristic reading is explicitly endorsed[493]. This is made clear in the introductory document provided with the 'levels' table:

> '[this hierarchy is] *a short-cut for busy clinicians, researchers, or patients to find the likely best evidence*. To illustrate you may find the following analogy useful. Imagine making a decision about treatment benefits in 'real time' (a few minutes, or at most a few hours). There are five boxes each containing a different type of evidence: which box would you open first? … we begin by

[489] (G. H. Guyatt & Drummond Rennie, 2002) p. 17
[490] (G. H. Guyatt & Drummond Rennie, 2002) p. 26
[491] (G. H. Guyatt & Drummond Rennie, 2002) p. 38
[492] (OCEBM Levels of Evidence Working Group, 2011)
[493] See also: (Howick, Chalmers, Glasziou, Trish Greenhalgh, Heneghan, Liberati, Moschetti, Phillips, & Thornton, 2011a, 2011b)

> searching for systematic reviews of randomized trials. If we didn't find any evidence in the systematic review box, you would go onto search for individual randomized trials, and so on[494]'

I suggest therefore that this contextualisation of supposedly 'strong' categorical statements further demonstrates that an epistemic reading of evidence hierarchies is mistaken. The shift to viewing hierarchies as making empirical claims; as advocated by Worrall and Howick, supports the view that hierarchies should be given a heuristic, not an epistemic, reading. Importantly, for the heuristic to be useful, one need not endorse any strong epistemological claims (such as the Categorical Interpretation). The heuristic interpretation only commits one to the claim that (if for example, one is interested in treatment benefits, then) randomised studies are most likely to offer stronger evidential support than non-randomised studies for a given hypothesis. If the empirical claim about the moderate effect size of most treatments is correct, then that commitment is met.

7.4 Summary

The Categorical Interpretation of EBM is not defensible; it falls foul of three well known problems described in §7.1: first, that the mere fact that a given research design was employed does not thereby entail that some particular study is good evidence for the hypothesis in question (the Bad Implementation problem). Even in the ideal case however there are problems: Second, that low ranked research designs are capable of providing good evidence, if the effect size is large enough (Dramatic Effects problem). Third, that highly ranked research designs may fail to give good evidence if the effect size is small enough (Small Effects problem).

This chapter described one way of solving these problems, proposed by John Worrall and Jeremy Howick. Their argument, which other authors have put in slightly different terms, is simply that good evidence should discriminate between plausible alternative hypotheses. The important point is that there is no a priori constraint on which research designs are capable of providing adequately discriminating evidence. Consequently giving evidence hierarchies – what I call – an 'epistemic reading' is

[494] (Howick, Chalmers, Glasziou, Trish Greenhalgh, Heneghan, Liberati, Moschetti, Phillips, & Thornton, 2011b) p. 1 [original emphasis]

mistaken. I suggest that evidence hierarchies should be read heuristically, and that the epistemological details of EBM are to be found in the techniques of critical appraisal. Indeed, I claim that critically appraising evidence just is the operationalisation of Worrall and Howick's account of evidence.

CHAPTER 8

8. Summary of Part Two

EBM is a difficult concept to rely on. Chapters 5 and 6 demonstrate that its details are unexpectedly unclear. The fact that, at the most general level, it is intuitively compelling makes it easy for arguments to become rhetorical. I claim that this is the explanation for the confusing picture of EBM in the medical literature, seen in Chapter 6.

The work here does not challenge the view that randomised trials often provide the best evidential support for medical claims. It does challenge the naïve view that they always and only do so. That view is worth challenging because it is the view which critics of homeopathy seem to hold (see Part One). Furthermore the work here challenges the view that there is a single coherent account of EBM in the medical literature. That result was unexpected, since many authors talk about progress and evolution of EBM.

In answer to the questions posed at the beginning of this part of the thesis, I claim that we should conclude that: (1) EBM, as put forward in the medical literature, does not provide a strong foundation for the evidential debate about homeopathy, but (2) that Worrall and Howick's account of how should EBM be interpreted provides a better foundation; with the consequence however that some re-evaluation is needed of the arguments put forward by opponents of homeopathy.

Proponents of homeopathy are quite right to criticise the Categorical Interpretation of evidence hierarchies. Whether they can conclude anything which is to their advantage however, is a further question. While proponents of homeopathy have legitimate objections to naïve formulations of EBM, it is an open question whether there are objections to more sophisticated formulations, such as suggested by Worrall and Howick.

One consequence of the view put forward by Worrall and Howick is that any kind of evidence may potentially offer support to a hypothesis. An evaluation of the evidence-base for homeopathy must be a critical appraisal of the total evidence. As a result, and contrary to the STC, it would appear to be less obvious that the best evidence for whether or not homeopathic treatment is a placebo comes solely from placebo-controlled trials. There are two points to consider:

First, the Implausibility Argument noted in Part One and above in Chapter 5, puts forward mechanistic evidence in support of the claim that homeopathic treatments are placebo treatments. On a Categorical Interpretation of evidence hierarchies, it would seem that this evidence can only ever offer weak evidential support for such a claim. Indeed, this is the position that the House of Commons' Science and Technology Committee take. On Worrall and Howick's view however, the Implausibility Argument may have a greater evidential role to play, since there is no a priori restriction on what kind of evidence can provide good evidence for a hypothesis. The conclusions of Chapter 7 therefore suggest that the reasoning involved in the Implausibility Argument deserves further attention: it may potentially weigh alongside evidence from clinical research. This will be examined in Part Four.

Second, clinical research – even placebo-controlled trials – is not likely to be decisive in the homeopathy controversy. The fact that homeopathic treatment, even if it is effective, does not have a large effect size, suggests that there will always be legitimate methodological reasons why the debate about the randomised trial evidence can be kept open. Calls for 'definitive' studies are naïve[495]. Calls for further research – further randomised trials – are problematic not simply because, as some have claimed, the answer is already known, but also because they are unlikely to settle the question.

8.1 Introduction to Part Three

In Part One it was shown that the Canonical Criticism uses placebos as the evidential standard that homeopathic treatments must surpass, if they are to be considered to work. The reasoning behind this relied on the idea that any medical treatment, if it is equivalent to placebo, and therefore inefficacious, cannot be said to "really" work. Outperforming placebo is the benchmark for therapeutic legitimacy in the Canonical Criticism. The normative role of placebos is also clearly evident from their use in the ethical arguments deployed in the policy debate, about homeopathy. Placebos are not just an evidential, but also an ethical, standard.

Part Three examines these two roles that placebos play. The questions to be addressed are as follows: first, what is the significance of placebo comparison? – the Canonical Criticism has a straightforward answer here, placebo comparison is

[495] (Baum & Edzard Ernst, 2009; Oberbaum et al., 2005)

significant because it, unlike other kinds of comparison, is best placed to distinguish between efficacy and effectiveness. Part Three examines in more detail whether that view can be sustained and what the distinction between efficacy and effectiveness amounts to. Second, why are placebo treatments though to be unethical? – Again, the Canonical Criticism has a straightforward answer, given by the No Placebos and Indirect Harm arguments in Part One. Namely, that giving placebo treatments is unethical. Part Three also examines this line of reasoning further.

To prefigure Part Three: the argument will be that examination of research into 'placebo effects' (Chapter 9) and reflection on the logic of placebo comparison (Chapter 10) shows that the way placebos are conceptualised in the Canonical Criticism cannot be sustained. The argument of Part Three shows that important revisions are necessary in the way that the evidential and ethical debates about homeopathy are framed and intertwined (Chapter 11).

PART THREE: PLACEBOS

CHAPTER 9

9. Placebos and the homeopathy controversy

According to the Canonical Criticism the key question that the evidential debate about homeopathy asks is whether homeopathic medicines *themselves* offer any therapeutic benefit; or, in other words, whether homeopathic treatments are efficacious. A distinction between efficacy and effectiveness, which is central to the Canonical Criticism, is drawn in order to highlight the fact that homeopathic treatment, like any medical treatment, may be or may appear to be effective for reasons other than the therapeutic effect of their constituents. Example scenarios are not difficult to imagine: a treatment may appear to be effective because an improvement in the condition, which was always going to occur, happens to occur at the same time as the treatment is given[496]. Or equally, a treatment may be effective because the condition improves as a result of other kinds of therapeutic effect, other than any effect of the treatment per se[497].

Ruling out these other explanations of a treatment's apparent efficacy provides reasons to believe that the treatment is genuinely efficacious. Fundamentally these attributions of efficacy are about being able to make causal claims[498], namely that a certain component of a treatment caused the therapeutic effects observed. In the case of homeopathy, according to the Canonical Criticism, determining whether homeopathic treatments themselves cause any therapeutic benefit necessitates the use of a special kind of experimental control: the placebo control[499]. Indeed placebo controls are frequently taken to be the *sine qua non* of efficacy testing. Ted Kaptchuk captures the point succinctly:

> 'Demonstrations of efficacy beyond placebo control in RCTs are fundamental to biomedicine's claim that its treatments are based

[496] Perhaps the condition has run its course (natural course of the disease), or perhaps the state of the condition is returning to more normal levels (regression to the mean).

[497] Or, a further alternative may be that the condition does not in fact improve, but the clinician believes it has. Perhaps because the patient says that it has, believing that is what they should say given that they have just received treatment. This would be a kind of measurement error.

[498] (Cartwright, 2007) (Cartwright, 2011a)

[499] It may be worth noting that a placebo control is not limited to use in randomised studies. For example, a placebo controlled study where patients are matched, rather than randomly assigned, to the treatment groups is perfectly possible to design. See: (Benedetti, 2009) pp. 9-12.

> on the objective physical–mechanical effects of pharmacology or physiological procedures and are not 'merely' rituals devoid of active ingredients. Placebo controls demarcate legitimate from illegitimate healing[500]'

The epistemic goal that motivates the use of a placebo control is ruling out precisely those therapeutic effects that are not related to the activity of the essential features of the treatment itself. Leaving discussion of the mechanistic argument in the Canonical Criticism aside (this will be revisited in Part Four), the contention in the Canonical Criticism is that homeopathic medicines cannot be considered to work unless they are shown to be more effective than a placebo in a randomised trial. Outperforming placebo is the benchmark for 'legitimate healing'. Reiterating Kaptchuk's point, Anne Harrington makes a similar observation:

> '[equivalence to placebo constitutes] a kiss of death for any therapy… To say that homeopathy (for example) gains its efficacy through the placebo effect is to say that it does not "really" work at all[501]'

The question to be addressed in this Part concerns what it is about placebo controls that makes them appropriate (or not) to provide the evidential and ethical standard for whether a treatment works.

The argument put forward in Part Three will be that evidence from contemporary placebo research, and reflection on the logic of placebo comparison itself, supports a substantial revision in how one thinks about the significance of placebo controls. To anticipate the conclusion: I claim that the term 'placebo' should be abandoned altogether; furthermore I claim that this has important consequences for views about the ethical provision of effective treatments more generally.

In §9.1 I review the experimental literature investigating placebo effects, and dismiss some intuitive explanations of them. In §9.2 I argue that the concept of placebos should be abandoned and in §9.3 I further examine the implications of those arguments.

---

[500] (Kaptchuk, 2011) p. 1849 See also: (Sullivan, 1993)
[501] (Harrington, 2002) p. 36 See also: (Harrington, 1997)(Wahlberg, 2008)

9.1 Placebos and placebo effects

A potential barrier to answering the question above might be problems surrounding the proper definition of a placebo and placebo effects. There have been many attempts to define what does and does not constitute a placebo. The problems such accounts face are significant[502]. It is important therefore to discuss a number of the puzzles around the conceptualisation of placebos, especially since some intuitive ideas one might have about placebos do not stand up in light of recent empirical investigations of placebo phenomena.

Clinical experiments investigating placebo phenomena have generated results that can appear to be unintuitive. Perhaps the paradigm case of a placebo effect is where sugar pills make one better (relieves pain, say) even though – so it seems – they shouldn't[503]. One part of making sense of placebo phenomena is to explain how sugar pills have this effect.

Indeed there are some intuitive ways that placebo phenomena like this have been explained. First, they have been explained in terms of subjective psychology – one merely feels better by taking the sugar pill[504]. Second, placebo phenomena have also been seen (perhaps melodramatically) as anomalies in the biomedical paradigm. They have been explained in terms of 'biomedical faith', or in terms of the 'irrationality' of biomedical practice, which speaks against the purportedly scientific status of biomedicine[505].

Both explanations receive support from the, again intuitive, view that placebos are inert substances incapable of producing therapeutic effects. Any effects they do seem to have are only apparent, and more strongly are 'fraudulent, deceptive, corrosive of medical authority, and therefore to be avoided[506]'.

---

[502] See for example: (A Grünbaum, 1981; A Grünbaum, Cicchetti, & Grove, 1991; Gøtzsche, 1994; Nunn, 2009a)

[503] Not all improvements in a condition, which are not due to active therapy, are therefore due to placebo effects. Other explanations of improvement, besides placebo effects include the condition running its course (natural course of the disease), the state of the condition is returning to more normal levels (regression to the mean) or other systematic effects from, for example, some efficacious parallel treatment; see: (Edzard Ernst & Resch, 1995; F. G. Miller & Kaptchuk, 2008; D. D. Price, Finniss, & Benedetti, 2008). Furthermore where placebo effects are positively harmful they are commonly referred to as *nocebo* responses (Barsky, Saintfort, Rogers, & Borus, 2002; Hahn & Harrington, 1997; Hahn & Kleinman, 1983) (S. R. Adler, 2011)

[504] (A. K. Shapiro & E. Shapiro, 1999) (Moerman, 2002a)

[505] (Comaroff, 1976; L. Price, 1984)

[506] (Moerman, 2002a) p. 400

The appeal of the first type of explanation lies in the temptation to see placebo phenomena in terms of psychological effects. The sugar pills, because they are inert, cannot not actually make one better, but only make one *feel* better. The example of a sugar pill apparently causing pain relief tempts one to produce a psychologising explanation, because it is not clear what other causal story to tell. The second type of explanation on the other hand suggests that placebo phenomena are a demonstration of the failure of (or at least points to gaps in) the biomedical paradigm to comprehensively explain the nature of healing. The supposed problem is that a therapeutic effect has been generated by a pill which, because of its inertness, lacked the capability (according to biomedical resources, so the argument goes) to produce such an effect. This is therefore taken to speak against the adequacy of those resources.

Both types of explanation represent two important ways that placebo phenomena are ordinarily conceptualised. They are explanations which understand placebo effects to be either 'in your head' or to demonstrate the 'limits of biomedicine'. The significance of results from contemporary clinical experiments that investigate placebo phenomena, which will be reviewed below, is that they show that neither of these explanations is adequate. Placebo effects are not just in your head, and the task of understanding placebo effects is not beyond the remit of biomedicine.

Three intuitive points about placebo effects have been noted; concerning two ways they might be explained and a further point about a common premise that both explanations share, regarding their supposed inertness. How these stand up against the contemporary literature will now be discussed in more detail. Firstly I argue that psychologising explanations are inadequate, second that placebo effects do not challenge the 'biomedical paradigm' and third clarify the sense in which placebos are 'inert'.

9.1.1 Psychologising explanations of placebo effects are inadequate

One puzzling result, and one which will be returned to, is – what can be called – the *naloxone result*. The experiment that first generated this result was conducted by Levine at al[507], however it has been noted that the study had some

[507] (Jon D Levine, Newton C Gordon, & Howard L Fields, 1978)

methodological flaws[508]. Never the less similar studies have been conducted and the result is well confirmed[509]. The experiment concerns the reduced ability of a placebo to alleviate pain, in combination with the drug naloxone. Naloxone is an opioid antagonist; which is to say, it inhibits the pain-relieving effect of opioids. Hence administration of the opioid painkiller morphine along with naloxone could be calibrated to produce little if any pain relief[510].

The logic of the experiment that generates the naloxone result is straightforward: patients (suffering from either clinical pain or experimentally induced pain [511]) are divided into four groups; following the 'open-hidden paradigm[512]'. Firstly patients are divided into two groups according to whether they are to receive an 'open' or 'hidden' treatment with saline, meaning that they are divided according to whether they know they are being given a treatment (and which they believe is a painkiller). Second, half of each of the two groups receive a dose of the drug naloxone (thus there are four different treatment groups). The different reactions of patients in these four groups are then compared.

The results go as follows: it was found that there was no difference in pain[513] when the two groups that received the hidden injection[514] were compared. That is to say, the presence or absence of naloxone had no effect on patients' pain when the saline injection was hidden. Which, of course, is what would be expected because saline injections do not contain opioids. Similarly as might be expected, it was found that the group receiving the open injection, but not naloxone, experienced a placebo response that reduced pain, compared to the two hidden groups. Fabrizio Benedetti

[508] (Grevert & Goldstein, 1978)
[509] See especially: (ter Riet, A. J. M. de Craen, A. de Boer, & Kessels, 1998)
[510] As we will see, naloxone can also eliminate analgesic placebo effects – but it is worth noting now that relatively large doses of naloxone are needed to eliminate these analgesic placebo effects, which gives us information about the type of opioid receptors involved, see: (Benedetti, 2009) p. 37
[511] In the studies reviewed by (ter Riet et al., 1998)
[512] For further explanation see, for example: (Finniss & Benedetti, 2005) and (Benedetti, 2009) p. 246-50.
[513] In most studies, pain is typically measured on a visual analogue scale (VAS), where patients place a mark on a 10cm strip, where the leftmost point equates to no pain and rightmost to the worst pain. Distance in millimetres from the left therefore provides a measure of patient's pain.
[514] The hidden injection is delivered by an automated infusion machine – no one is present, and the patient is unaware of the infusion.

states: 'telling the patient that a painkiller was being injected (with what was actually a saline solution) is as potent as 6-8mg of morphine[515]'.

Hence, seeing that one is having an injection and being told that it is a powerful painkiller generates a placebo effect. Comparing the two open groups however reveals a more counter intuitive result. The open group which did not receive the naloxone experienced greater pain relief than the open plus naloxone group. This is to say: the presence of naloxone inhibited the placebo effect.

Since saline injections certainly do not contain opioids, this result is a demonstration of the fact that placebo effects can be mediated through physiological mechanisms. The open injection and assurance that it is a painkiller has a demonstrable physiological effect; specifically, it mobilises the patients' endogenous opioids[516]. In the hidden groups this was not the case. In the open group that received naloxone, the action of the patients' endogenous opioids was blocked. Consequently, this explains why only the open group which did not receive naloxone experienced a placebo effect.

This result receives support from a further type of experiment. It has also been shown that if saline solution is given (again on the promise that it is a painkiller) in combination with proglumide (which enhances opioid mediated responses[517]), then greater analgesic effects are reported[518]. If however the saline is given in 'hidden' rather than 'open' conditions (as explained above), then no differential effect between saline alone and saline in combination with either naloxone or proglumide is observed[519].

The naloxone result, and the closely related proglumide result, provides strong evidence that some painkilling placebo effects are mediated by patients' endogenous opioid systems. Opioid antagonists diminish the painkilling placebo effect, opioid agonists increase it. Moreover a third type of experiment corroborates this picture. In studies of patients with chronic pain, those who responded best to placebo treatment have been found to have a higher level of endorphins (which are

[515] (Benedetti, 2009) p. 71 – referring to the study by (J D Levine, N C Gordon, R Smith, & H L Fields, 1981)
[516] (Benedetti, 2009; ter Riet et al., 1998)
[517] In effect proglumide is an opioid agonist; albeit by an indirect mechanism (it is a cholecystokinin antagonist) (Benedetti & Amanzio, 1997) (Benedetti, 2009) p. 75
[518] (Benedetti & Amanzio, 1997; Benedetti, Amanzio, & Maggi, 1995; ter Riet et al., 1998)
[519] (ter Riet et al., 1998)

endogenous opiates) in their cerebrospinal fluid, moreover, the painkilling placebo effect and higher levels of endorphins could be blocked by naloxone[520].

Interestingly however not all painkilling placebo effects can be eliminated by naloxone, or enhanced by proglumide. For instance, Amanzio and Benedetti investigated placebo analgesia by treating patients with ketorolac (a non-opioid analgesic) for two days, then switching them to placebo[521]. In this case, the analgesic effect was maintained even in combination with naloxone. This suggests that the mechanism through which a painkilling placebo effect acts can be conditioned by prior contact with particular kinds of painkiller[522]. In short, there is no single mechanism for placebo analgesia, or therefore for placebo effects in general[523].

Another result, similar to the naloxone result, which is of particular note is – what can be called – the *carisoprodol result*. This refers to a number of related studies about the ability of placebos to modify drug responses[524], but most particularly to the study by Flaten et al[525]. As with the naloxone result, the logic and results are straightforward:

In the study by Flaten et al healthy subjects were divided into three groups, depending on whether they were given either no information about the drugs they would receive, were told they would receive a stimulant, or were told they would receive a relaxant. After all patients were forewarned that they may receive either an active drug or placebo, members of each of the three groups were given (without their awareness) either lactose capsules or capsules containing the drug carisoprodol. Carisoprodol is a centrally-acting skeletal muscle relaxant[526]. The different responses of the six groups were then compared.

As might be expected in the groups that did not receive carisoprodol, the information that they had received a stimulant increased the subject's tension[527] compared to the other non-carisoprodol-receiving groups. Furthermore those other two non-carisoprodol-receiving groups (that is, the no-information and the relaxant-

[520] (D. D. Price et al., 2008)
[521] (Amanzio & Benedetti, 1999)
[522] (Benedetti & Amanzio, 1997; Benedetti et al., 2003)
[523] (L Colloca & Benedetti, 2005; D. D. Price et al., 2008)
[524] (Flaten, 1998, 2009; Flaten, Simonsen, & Olsen, 1999; Flaten et al., 2004)
[525] (Flaten et al., 1999)
[526] Meaning that it acts on the central nervous system, as opposed to neuromuscular junctions. The relevant point however is merely that it is a muscle relaxant.
[527] Tension was measured by eight different reflex test, including, for example skin conductance and blink response. See: (Flaten et al., 1999) pp. 251-3

information groups) both experienced decreases in tension. The two groups given no information about what they were receiving experienced highly variable effects, regardless of whether they also received carisoprodol or not.

These results accord with our intuitive expectations regarding placebo effects. The most interesting findings however, concern the group which received both carisoprodol and the information they were being given a stimulant. In this group, the subjects' tension increased in comparison to the group given the same information, but no carisoprodol[528]. This is to say: the presence of carisoprodol had the opposite effect from what would be expected. Instead of relaxing the subjects, its presence augmented the verbal-stimulant effect. Additionally it was found that levels of carisoprodol in the blood were higher in the group told they were receiving a relaxant, compared to group told they were receiving a stimulant. The belief that they were given a stimulant slowed down their absorption of carisoprodol.

These results – concerning naloxone, proglumide, keterolac, and carisoprodol – each demonstrate specific physiological mechanisms behind placebo effects, despite the fact that the placebos used contain no substance that plays a role in the mechanisms by which those effects are generated. The placebo effect, in the case of the naloxone and proglumide results, is mediated by endogenous opioid systems and observable through patient reported outcomes as well as neuroimaging techniques[529]. In the case of the keterolac result the same analgesic effect is produced, but without enrolling patients' endogenous opioids. In the case of the carisoprodol result the expectation of a stimulant effect changed the absorption rate of carisoprodol and was, counter-intuitively, augmented by the presence of carisoprodol. The key point which each of these results illustrate is that the 'placebo effect' is more than a psychological phenomenon. Indeed placebo effects of various sorts, relating to Parkinson's and Alzheimer's as well as pain, have been shown to be mediated by a variety of physiological mechanisms[530].

This empirical evidence suggests therefore that psychologising explanations are simply not an adequate way to describe placebo effects. Such explanations fail to do justice to placebo phenomena in two ways: firstly by failing to capture the fact that placebo effects operate in specific ways, through physiological mechanisms; and

[528] (Flaten et al., 1999)
[529] (D. D. Price et al., 2008; ter Riet et al., 1998)
[530] (Benedetti, 2009; Koshi & Short, 2007; F. G. Miller & Kaptchuk, 2008; D. D. Price et al., 2008; Stewart-williams & Podd, 2004)

secondly by failing to capture the range of different mechanisms through which placebo effects operate[531].

9.1.2 Placebo effects do not demonstrate the 'limits of the biomedical paradigm'

The contemporary research literature provides evidence that placebo effects are not, in general, merely psychological effects. The results presented above show various instances where placebo effects have physiological effects in a precise and specific sense. Furthermore these results demonstrate interplay between psychology and physiology. Never the less the relative ignorance of the underlying mechanisms and subjective psychology might be thought to be a plausible basis on which to claim that medical science is ill-equipped (perhaps in principle) to adequately explain placebo effects. Like psychologising explanations of placebo effects however, this line of thought is not supported by evidence from the contemporary research literature.

Linnie Price provides one of the few sociologically-oriented analyses of the implications of placebo effects[532]. It is worth looking at her argument in more detail. Whilst Price's paper is over twenty years old, it is notable that her arguments have much in common with some of the recent arguments put forward by proponents of homeopathy. There are similarities, for example, to the idea that homeopathic treatments are too 'complex' to be investigated in placebo controlled trials (see Part One). These arguments will be returned to in Part Four.

Price claims that 'the implication of the placebo effect for medicine… is that it relocates healing in the realm of the irrational[533]' and moreover she states that:

> 'to accept the implications of the placebo effect would be to challenge the claims to truth of *all* medical knowledge: it would necessitate a paradigmatic revolution of untold proportions[534]'.

[531] Although it must be pointed out that the extent of the generalisations that can be based on results from studies of pain has been questioned. For instance, Arthur and Elaine Shapiro argue against assuming a wide significance for the results of placebo analgesia studies (A. K. Shapiro & E. Shapiro, 1997a) p. 232
[532] (L. Price, 1984) See also: (Comaroff, 1976)
[533] (L. Price, 1984) p. 71
[534] (L. Price, 1984) p. 69

Her argument for these conclusions rests on the idea that the biomedical paradigm is necessarily antithetical to the social character of placebos effects, because the range of contextual and socio-cultural factors that contribute to placebo effects are not amenable to scientific study[535]. Price also makes a further argument, which trades on an air of self-reflective paradox. She argues that the image of 'scientificity' that medicine cultivates around itself – a kind of biomedical faith - may be what is responsible for generating placebo effects; and the existence of which undermines that very notion of scientificity.

Both of Price's arguments are premised on an incommensurability claim. She asserts that the 'reductive' biomedical paradigm cannot accommodate placebo effects, because those effects are partly and irreducibly constituted by social meaning[536]. No clear argument is offered for this assertion[537].

Even granting the claim about the scientificity of medicine, Price's argument ignores the fact that the factors which contribute to placebo responses are heterogeneous[538]. Price is right to claim that the perception of medicine as 'scientific', and faith in the power of medicine, are indeed likely to be factors that contribute to placebo effects; but, as will be described below and as the experimental results described above show, contemporary research is able to both unpack the different elements which are captured by Price's notion of scientificity, and in doing so, refute the charge that it undermines itself.

More importantly, in contrast to Price's argument. The naloxone and carisoprodol results show that empirical investigation of placebo effects is both possible and able to take account of patients' perceptions and the context of treatments, scientifically. This is precisely what is done in experiments that utilise open and hidden treatments, or control the information that is given to patients. Moreover much of the research that will be introduced below shows that a similar level of sophistication exists in relation to the investigation of the many different contextual factors that influence placebo effects. Placebo effects can and have been

[535] (L. Price, 1984) pp. 67-9
[536] (L. Price, 1984) pp. 67-9
[537] It seems more probable that the implications of placebo phenomena, for issues of biological reductionism, are rather minimal. It is simply not clear what implications the existence of placebo effects have for these wider metaphysical debates. More significantly, debates about the metaphysics of the biomedical paradigm seem straightforwardly irrelevant to the issue here – that is, whether placebo effects can be meaningfully investigated.
[538] (Koshi & Short, 2007; D. D. Price et al., 2008; Stewart-williams & Podd, 2004)

studied in depth, using the resources of biomedical science. As it stands, an argument such as Price's is too hasty in drawing its conclusions. Assertions that there are 'social' components of placebo effects do not entail that those components cannot be studied empirically, or that they are beyond the limits of biomedical science.

9.1.3 Confusion about placebos' supposed 'inertness'

The thought that placebos both are inert and yet cause placebo effects has resulted in much confusion. The following is (part of) an influential definition put forward by Arthur and Elaine Shapiro[539]:

> '[A placebo] is any therapy prescribed knowingly or unknowingly by a healer, or used by laymen, for its therapeutic effect on a symptom or disease, but which actually is ineffective or not specifically effective for the symptom or disorder being treated[540]'

The definition is framed in terms of inert substances (and sham procedures could be included) and their non-specific effects. The definition is cast in causal language and it invites one to conceive of an 'inert' substance, a placebo, endowed with causal powers to bring about non-specific effects. Whilst perhaps this is an intuitive picture of placebos and placebo effects, many authors have argued that this is at best a very confusing picture[541,542].

It should be noted that no one argues that placebos are inert in any absolute sense[543] (indeed no substance is completely inert[544]). They are inert only in the sense that they are not effective for the condition being treated. Contrary to the idea that placebos are absolutely inert, substances may in fact be deliberately used in a placebo controlled trial (PCT) precisely because of their ability to induce effects similar to the side-effects of standard treatments. For example, a number of trials of anti-depressants have included so-called 'active' placebos. In some trials of anti-depressants atropine is included as a constituent of the pills given to placebo

[539] See varieties of it in: (A. K. Shapiro, 1964, 1968; A. K. Shapiro & E. Shapiro, 1997a)
[540] (A. K. Shapiro & E. Shapiro, 1997b) p. 12
[541] (Caspi & Bootzin, 2002; A Grünbaum, 1981; A Grünbaum et al., 1991; Gøtzsche, 1994, 1995; Lichtenberg, Heresco-Levy, & Nitzan, 2004; F. G. Miller & Kaptchuk, 2008; Moerman, 2002b; Nunn, 2009a) (F. G. Miller & H. Brody, 2011)
[542] See also Chapter 9, below.
[543] (A. K. Shapiro & E. Shapiro, 1997a, 1999)
[544] (Howick, 2011) p. 81

groups[545]. The purpose of this is to mimic some of the known side-effects of anti-depressant treatments; such as experiencing a dry-mouth[546]. By maintaining a consistent experience between treatment groups, this reduces the likelihood that patients or clinicians become unblinded, and therefore also reduces the likelihood that patients experience an increased or decreased therapeutic effect on account of their knowledge of which group of the trial they are allocated to.

Of course, the validity of 'active placebo' research designs such as this would be questionable if – to take the above example again – atropine itself was known to have depressive or anti-depressive effects because in that case the 'active' placebo would fail to be inert in the required sense[547]. Given that atropine does not have depressive or anti-depressive effects it can legitimately be used as a control in the placebo group: the tablet containing the atropine is inert, in the required sense[548].

The key point about placebos' 'inertness' is that the placebo itself must not have an effect on the patients' condition. There is no absolutely inert substance, but that it not what is required. Importantly there may be situations where it is useful for placebos to have particular effects. This is what is meant by the claim that placebos are 'inert'. The obvious consequence of the inertness of placebos is that it rules them out of any straightforward role in the causal explanation of placebo effects[549]. Daniel Moerman and Wayne Jonas therefore counter-intuitively claim: 'the one thing of which we can be absolutely certain is that placebos *do not* cause placebo effects[550]'.

[545] For examples of such trials see especially: (J Moncrieff, Wessely, & Hardy, 2004) See also: (Mora, Nestoriuc, & Rief, 2011)
[546] (Joanna Moncrieff, 2009) p. 147
[547] It seems this is not well investigated, see: (J Moncrieff et al., 2004) p. 26
[548] More generally, the inclusion of any agents in the control group which have a therapeutic effect on the condition being investigated is a potential threat to a trial's validity. This is a potential problem for any placebo controlled trial, not just those which include 'active' placebos. Consequently Beatrice Golomb et al have recently called for better reporting of the constituents of placebos used in clinical research. See (Golomb et al., 2010)(Golomb, 1995). They cite the example of a number of trials of cholesterol reducing drugs, for the treatment of cardiovascular risk. The placebos used contained olive oil or corn oil, which were used as excipients, but which are also known to have cholesterol lowering properties. Beatrice Golomb noted in 1995 that: '[the FDA] sets no regulations on the constituents of placebos… no systematic efforts are made to ensure the inertness of placebos'. A more systematic investigation of the reporting of placebos by Golomb et al in 2010 revealed the picture had not changed.
[549] (F. G. Miller & Kaptchuk, 2008; Moerman, 2002b; Moerman & W. B. Jonas, 2002; D. D. Price et al., 2008)
[550] (Moerman & W. B. Jonas, 2002) p. 471 original emphasis

The next section addresses the question of what does cause placebo effects, if not placebos?

9.2 Which components of a treatment can cause placebo effects?

In what follows the factors which have been shown to cause placebo effects will be described in more detail, along with a description of the variability of placebo effects and, what has been called, the 'Additivity assumption'. They key point to emphasise is that there is a diverse range of treatment components have been shown to contribute to placebo effects.

9.2.1 A diverse range of components

Many different components of a treatment have been shown to generate placebo effects[551]. The main way that these different components have been conceptualised will be introduced. To anticipate the discussion, the most prominent way to unify the ability of many different components of a treatment to generate therapeutic effects is by reference to the *meaning* that is attached to them. Ayo Walhberg has argued that the work of medical anthropologists has done much to 'decriminalise' the notion of placebo effects in recent decades, by providing a (biomedically acceptable) link between the symbolic and the physiological[552]. Indeed the work of medical anthropologists has been crucial for understanding the particular patterns of meaning and symbolism that are involved in placebo effects[553] (see §9.3 below). Initially however, four different types of components of treatments, which have been shown to contribute to placebo effects, will be described.

First there are idiosyncratic components of the therapeutic context; including such factors as particular verbal suggestions[554], as well as the attitude, enthusiasm and behaviour of the healthcare team. More generally these are components of a treatment that pick out all those verbal and non-verbal ways that patients and

[551] In relation to placebo analgesia, see for example: (Vase, Nørskov, Petersen, & D. D. Price, 2011) pp. 1914-9
[552] (Wahlberg, 2008)
[553] (S. R. Adler, 2011; Csordas, 1988; Kleinman, 1977; Kleinman, Mendelsohn, & Elkana, 1981; Scheper-Hughes & Lock, 1987; J. J. Thompson, Ritenbaugh, & Nichter, 2009)
[554] Such as the suggestion 'This has been shown to reduce some people's pain' – not quite a lie, even in the case of saline solution or sugar pills; but still ethically suspect? See (Lichtenberg et al., 2004)

clinicians interact to create a given treatment context. Many such aspects have been shown to generate placebo effects[555]. As Colloca and Miller state:

> 'The doctor promotes healing by means of clinical attention (including the ritualistic element of administering treatment) and communicative interaction with the patient, including reassurance, verbal suggestions for positive therapeutic expectations, empathic listening, and encouragement[556],

Second there are more tangible contextual components. For example, the means by which a treatment is delivered[557]; as demonstrated, for example, by the 'open-hidden' experiments described above, where painkillers delivered by a hidden mechanical pump were shown to have less of an analgesic effect compared to delivery of the same painkiller by an 'open' method, the patient was aware of[558]. Again Colloca and Miller pick out such factors as:

> 'the therapist's white coat, diagnostic instruments, the appearance of the therapist's office or hospital room… the vehicles of treatment (e.g. syringe or tablets)[559],

Third there are the cognitive and emotional states of the patient. Both Price et al[560] and Stuart-Williams and Podd[561] emphasise the role that patients' expectations have been shown to play in generating placebo effects. Both cite a range of studies that paint a complex picture of the interaction between patients' expectations, avoidance and attainment desires, focus on particular goals as well as previous positive or negative experiences with particular treatments[562]. The carisoprodol result can usefully be seen in these terms, too. Consider those patients told they were receiving a stimulant, but in fact received carisoprodol. Patients were

[555] (H. M. Adler & Hammett, 1973) (Blasi, Harkness, Edzard Ernst, Georgiou, & Kleijnen, 2001; Kaptchuk, 2002; Ong, De Haes, Hoos, & Lammes, 1995; D. D. Price et al., 2008)
[556] (Luana Colloca & F. G. Miller, 2011) p. 1864
[557] (Amanzio, Pollo, Maggi, & Benedetti, 2001)
[558] (Benedetti, 2009)
[559] (Luana Colloca & F. G. Miller, 2011) pp. 1860-2
[560] (D. D. Price et al., 2008)
[561] (Stewart-williams & Podd, 2004)
[562] For example: (Kirsch, 1985)

told to expect a stimulant effect, and they experienced 'an effect' after receiving the (unknown to them) drug carisoprodol. The experience that *something* was affecting them reinforces the stimulant effect, which they expect to feel, despite the fact that the drug they have been given is a muscle relaxant. Of course what is astonishing about the carisoprodol result is that the reinforced perception that one is being stimulated not only diminishes the drugs effect, but reverses it.

Fourth there are those aspects of a treatment which tap into wider socio-cultural values and thereby generate placebo effects[563]. Surgery for instance, and 'physical' therapies generally, have been shown to induce placebo effects seemingly on account of their dramatic and visceral nature[564]. Two notable studies in this regard are Cobb et al[565] and Dimond et al[566], both of which looked at sham surgery for angina and found that patients who underwent the sham surgery showed the same levels of improvement as those that had the full procedure (mammary-artery ligation)[567].

Other socio-cultural factors, with a less dramatic perception than surgery, have also been shown to play a role in generating placebo responses[568]. Considering pills for instance, the number given[569], their colour[570] and their branding[571] and marketing[572] have all been shown to affect the magnitude of placebo responses. Blackwell et al demonstrated in a group of medical students – told to expect either a stimulant or sedative effect from the pills they were given – that two pills had more of an effect than one, and that blue pills had a more sedative effect than pink ones[573]. The results does not generalise unproblematically however, Moerman notes that the blue=sedative association responsible for the effect observed by Blackwell et al does

[563] (Kirmayer, 1993, 2004; Moerman, 1983, 2000, 2002a, 2002b; Moerman & W. B. Jonas, 2002; Moerman et al., 1979)
[564] (A. G. Johnson, 1994; Kaptchuk, P. Goldman, D. A. Stone, & W B Stason, 2000)(Moerman & W. B. Jonas, 2002)
[565] (Cobb, G. I. Thomas, Dillard, Merendino, & Bruce, 1959)
[566] (Dimond, Kittle, & Crockett, 1960)
[567] Moerman and Jonas note that as a consequence of these studies, the procedure is no longer performed (Moerman & W. B. Jonas, 2002)
[568] (Barrett et al., 2006)
[569] (Blackwell, Bloomfield, & Buncher, 1972; A. J. M. de Craen et al., 1999; Moerman, 2000)
[570] (Blackwell et al., 1972; Cattaneo, Lucchelli, & Filippucci, 1970; A. J. M. de Craen, Roos, De Vries, & Kleijnen, 1996; Schapira, McClelland, N. R. Griffiths, & D. J. Newell, 1970)
[571] (Branthwaite & Cooper, 1981)
[572] (Weissman et al., 2003)
[573] (Blackwell et al., 1972)

not apply, for instance, to populations of Italian men, apparently because blue is a highly charged and exciting colour: it is the colour of the national football team[574].

Consequently, the most interesting, and perhaps most coherent, approach to understanding placebo effects suggests that they should be conceived of as the result of a range of context-specific psychological and social factors, operating through specific physiological mechanisms[575].

9.2.2 Variability and magnitude

The magnitudes of placebo effects are highly variable. It is often quoted, from Henry Beecher's paper[576], that 30-40% of the effectiveness of a treatment can be attributed to placebo effects, however this misrepresents Beecher's original claim[577] and, moreover, Beecher's method does not distinguish between placebo effects and improvement that would have occurred anyway[578]. It is also inconsistent with recent results which show that, for example, in trials of new drugs for depression the magnitude of placebo effects has been increasing over time[579]. Additionally there is evidence to suggest that the variability of placebo effects is not explained by any particular psychological or demographic characteristics of patients, instead, as Daniel Moerman argues, the difference between placebo responders and non-responders is to be explained by: 'what patients know (not what kinds of people they are) and what things mean[580]'.

It should also be noted that placebo effects do not necessarily have a positive effect on patients. 'Nocebo' effects, that is, placebo effects with a negative effect on the patient, are common. For example, respiratory depression has been shown to follow administration of placebo analgesics (mediated by endogenous opioids) – as would be expected had opiates been given[581]. Also, side-effects normally induced by non-placebo interventions often occur in the placebo group of clinical trials[582]. In a

[574] (Moerman, 2002b)
[575] (Hahn & Kleinman, 1983; Kaptchuk, 2002; Kirmayer, 2004; Moerman, 2002b; Moerman & W. B. Jonas, 2002; Papakostas & Daras, 2001; L. Price, 1984; Stein, 1983)
[576] (Beecher, 1955)
[577] (Moerman, 1983) p. 13
[578] (Gøtzsche, 1994; D. D. Price et al., 2008)
[579] (Walsh, Seidman, Sysko, & Gould, 2002)
[580] (Moerman, 2002b) p. 46
[581] (D. D. Price et al., 2008)
[582] (Barsky et al., 2002; D. D. Price et al., 2008; L. Price, 1984)

recent placebo controlled trial of a four drug polypill, the authors note that the placebo group experienced high rates of side-effects of the sort that would be likely to be experienced were they receiving the drug treatment[583]. Most notably, patients in the placebo group experienced gastric irritation. This is a common side-effect of aspirin and which the patients knew was a component of the polypill. Indeed one might speculate that the fact that the trial population (individuals with low cardiovascular risk) had previous experience with the individual treatments contained in the polypill (and therefore will have known the kinds of side-effect they might experience) contributed to the high rates of side effects in the placebo group.

In addition to the variability of placebo effects, it has been argued by Asbjørn Hróbjartsson and Peter Gøtzsche that there is little evidence to support the claim that placebo effects are powerful effects at all[584]. Their key point is that there are few studies which compare placebo- with no-treatment[585] hence reading off the magnitude of the 'placebo effect' from the effects seen in placebo groups in clinical trials does not capture – what Ernst and Resch[586] call – the 'true placebo effect'. A further point to note is that studies which make the comparison between placebo and no-treatment groups are particularly susceptible to response bias because, for example, the fact of being treated is likely to be a significant influence on a patient's reporting of the state of their condition[587]. In their meta-analyses Hróbjartsson and Gøtzsche examine only clinical trial data which contain both a placebo and no-treatment group. Their conclusion – in the most up-to-date version of their analysis – is that placebo effects had a highly variable but significant effect on pain, and a modest effect on subjective outcome measures, but otherwise were 'small and uncertain[588]'.

Their results generated much debate[589] and are often questioned in the rest of the literature[590]. However Hróbjartsson and Gøtzsche's results are compatible with

[583] (PILL Collaborative Group, 2011)
[584] (Hróbjartsson & Gøtzsche, 2001, 2004) (Hróbjartsson & Gøtzsche, 2010)
[585] (Hróbjartsson & Gøtzsche, 2001, 2004) (Hróbjartsson & Gøtzsche, 2010) (Kienle & Kiene, 1997)
[586] (Edzard Ernst & Resch, 1995)
[587] See most notably: (Hróbjartsson, Kaptchuk, & F. G. Miller, 2011) – and note further the usual risks of publication bias etc.
[588] (Hróbjartsson & Gøtzsche, 2010) p. 17
[589] There conclusion in 2001, was slightly more negative. For examples of the response it generated, see the many letters in the *New England Journal of Medicine*: 2001, 345(17), pp. 1276-9
[590] See for example: (Stewart-williams & Podd, 2004) pp. 326-8

the view that placebo effects are highly variable, both with respect to particular conditions as well as across conditions[591]. However, the results speak against the view that placebo effects have a moderate and stable effect size, across and within many different conditions[592]. Of course, one might argue that this is precisely what one might expect to find if, as Moerman argues, the key to understanding placebo effects is a treatment's meaning to individual patients. The reason being that if placebo effects are sensitive to idiosyncrasies of patients' attribution of meaning, then they are less likely to be robust across very different circumstances.

It might therefore be argued that by pooling studies of different placebos and different conditions, Hróbjartsson and Gøtzsche are, in effect, mixing paint colours to get brown. However this objection must take account of the subgroup and meta-regression analyses that Hróbjartsson and Gøtzsche performed in order to explain the heterogeneity of the trials they analysed[593]. Slightly larger placebo effects were found when outcome measures were either patient reported or observer reported (but required some input from the patient), as well as when the studies explicitly set out to study the placebo effect. Placebo effects were also larger when the placebo used was 'physical[594]' and when the patients did not know they were receiving placebo[595]. As it stands Hróbjartsson and Gøtzsche's explanation of the heterogeneity of the trials they analysed is consistent with the effect of the kinds of factors outlined in §9.2.1, but it speaks against the average effect of those factors being large.

More fundamentally the objection that pooling is illegitimate ignores the fact that, if there truly were a large or moderate average effect size, then it ought to be detectable by their method. Recall for example the meta-analysis performed by

[591] Expressed for instance in: (Koshi & Short, 2007; Moerman, 1983, 2002b; D. D. Price et al., 2008)

[592] Which is the view expressed most famously in (Beecher, 1955).

[593] In this context, heterogeneity means whether the trials analysed were statistically dissimilar (as expressed by the I^2 statistic) that is, 'whether there are genuine differences underlying the results of the studies (heterogeneity), or whether the variation in findings is compatible with chance alone (homogeneity)' (Higgins, S. G. Thompson, Deeks, & Altman, 2003) p. 557 See also: (Perera & Heneghan, 2008) – in the case of Hróbjartsson & Gøtzsche's most recent meta-analysis, the I^2 for the overall pooled analysis was 45% (p. 306) for continuous outcomes and 42% (p. 312) for binary outcomes. These results are high enough to warrant the further investigation undertaken in the meta-analysis: (Hróbjartsson & Gøtzsche, 2010).

[594] Such as sham acupuncture or a switched-off medical device; as opposed to pharmaceutical (e.g. pills) or psychological (e.g. a neutral conversation) placebos. See (Hróbjartsson & Gøtzsche, 2010) p. 6

[595] (Hróbjartsson & Gøtzsche, 2010) p. 17 and 18

Shang et al[596], cited in Chapter 2. In that meta-analysis, trials of homeopathic medicine were matched with trials of conventional medicine for a range of different conditions; the result being that Shang et al did find a significant average effect of 'conventional medicine' in general.

9.2.3 Additivity

Additivity is the claim that the efficacy of different components of a treatment combine by adding their effects. Relatively little research has investigated the legitimacy of Additivity in general or in particular circumstances[597]. Additivity underlies the practice of subtracting the average outcomes in different treatments groups of a trial in order to isolate the differential effect of a treatment. Put in terms of drugs, Kirsch summarises a similar idea as follows: 'The additive assumption is that the effect of the drug is limited to the difference between the drug response and the placebo response[598]'. In general the truth of Additivity cannot be assumed. Interestingly, there are at least three ways that Additivity may fail:

Threshold effects. There are upper limits on both how fast a condition can be improved, and the extent to which it can be improved. Consider that for at least some conditions once one is 'better' extra treatment does not make one 'more better'. If placebo effects already have an impact on a condition, then even if the placebo effects and treatment effects are additive, the existence of this kind of threshold is likely to result in the difference between placebo and treatment underestimating the efficacy of the treatment. To make this clearer, consider an analogy: if one is free-wheeling downhill on a bicycle, then pedalling slowly does not increase one's speed further. Strictly this is not a failure of Additivity as such, but of the ability to adequately measure the additive effect (because there is a threshold).

Overdetermination effects. As well as being unable to measure the effect of some treatment, because there is an upper limit on how effective a treatment can be, one may also fail to measure the effect of some treatment because its

[596] (Shang, Huwiler-Müntener, et al., 2005)

[597] That Additivity is often, if not always, assumed has been noted by: (Enck, Klosterhalfen, Weimer, Horing, & Zipfel, 2011; Howick, 2011; Kaptchuk, 2001; Kirsch, 2000; Kleijnen, A. J. D. Craen, Everdingen, & Krol, 1994; Meissner, Kohls, & Luana Colloca, 2011)

[598] (Kirsch, 2000) p. 733 See also: (Enck et al., 2011) p. 1890 and (Meissner et al., 2011) pp. 1784-5 - Kirsch goes on to give some tentative reasons why additivity is true for anti-depressant medicines.

effectiveness is duplicated by placebo effects[599]. This is a true failure of Additivity. If placebo effects impact on a condition, and the placebo and treatment effects are not additive, then some diminishing of the placebo effects may not diminish the overall effect. If these different components were additive, then one would expect that a reduction in one component would result in a reduction overall. When Additivity fails, because of some degree of overdetermination, the treatment 'picks up the slack' because it duplicates some of the effectiveness of the placebo. Analogously, if one is pedalling redundantly whilst free-wheeling downhill, one can maintain a constant speed as the gradient flattens, as one's pedalling becomes non-redundant.

Interaction effects. The efficacy of one component of a treatment may be modulated by another. The most striking example of this is the carisoprodol result, reported by Flaten et al and discussed above in §9.1.1[600], but expectation-modulated drug responses have also been demonstrated, for example, with stimulants[601] and nicotine gum for smoking cessation[602] as well. Additivity fails here because the placebo effect impacts not only on the condition, but also on the effect of the treatment. The overall effect may be either under- or overestimated, depending on whether it is augmented or negated by other components of the treatment. To continue the cycling examples, one might draw an analogy with the fact that the peloton is able to maintain a higher average speed, and for longer, than an individual cyclist. In that case, the cyclists in the peloton interact synergistically to go faster.

As a number of authors have noted, there is a relative paucity of evidence for the legitimacy of Additivity[603]. Enck et al review the existing literature, and marshal clinical, mathematical and neurobiological evidence, concluding that 'the additive model is at question[604]'; Linde et al come to the same conclusion:

> 'the available studies suggest that context factors not only have direct effects but also interact with specific effects by either

[599] This model is discussed, and represented visually, in: (Kirsch, 2000) fig. 1 see also: (Enck et al., 2011) fig. 2
[600] (Flaten et al., 1999)
[601] (S. H. Mitchell, Laurent, & Wit, 1996)
[602] (Hughes, Gulliver, Amori, Mireanlt, & Fenwick, 1989)
[603] (Enck et al., 2011; Kaptchuk, 2001) (Kirsch, 2000) (Howick, 2011)
[604] (Enck et al., 2011) pp. 1891

> increasing or decreasing the differences between active treatment and placebo[605]'

It seems that one needs some justification to believe that Additivity does hold in a particular case. If Additivity is assumed to hold, then it at least ought to be stated whether there is, or is not, some evidence for such an assumption. As most of the authors referenced above note, the legitimacy of Additivity is rarely considered in clinical trials. The consequence of the, in general, failure of Additivity[606] will be considered below (see §10.2.2) and in Part Four. However note briefly that, if one considers the notion of a 'complex intervention' noted in Part One, then the failure of Additivity in general perhaps suggests that all treatments are, to some extent, complex interventions. Interesting cases arise when the magnitude of the interaction effects are sufficient to undermine the approximation to Additivity.

9.3 Meaning-theories of placebo effects

The quote in §9.1.3 from Moerman and Jonas – 'placebos do not cause placebo effects' – highlights the way in which placebos must be inert. If placebos did cause placebo effects, then they would not be inert in the required sense. Consequently however this might seem to make placebo effects even more puzzling. The question then is how should one conceptualise the counter-intuitive causal chains that seem to underlie, for example, the elimination by naloxone of the pain relieving effect of saline solution, or the augmentation by the relaxant carisoprodol of an expected stimulant effect? Again, the contemporary literature offers a coherent answer. As noted above, the literature offers an account that frames placebo effects in terms of meaning: call these 'meaning-theories'. The motivation for such theories is as follows:

Recent work has responded to the problem of making causal links between placebos and placebo effects by severing the need for such links. Firstly by identifying placebo effects with effects that result from the 'simulation of an active therapy

[605] (Linde, Fässler, & Meissner, 2011) p. 1909

[606] The phrase "in general" can sometimes seem ambiguous, so to be clear: by 'failure of Additivity, in general' I simply mean that Additivity does *not always* hold. I do not mean to imply that it *mostly does not* hold.

within a psychosocial context[607]'. Secondly, by restricting the term 'placebo' to denoting an often present, but singularly insufficient component of these effects; placebo effects do not require there to be a 'placebo', as such[608] (one doesn't think of the different treatment components listed in §9.2.1 as 'placebos' for example).

The most important insight from this work is the reconceptualising of 'placebo effects' in terms of the meaning and significance of the treatment and the treatment context. Such a change in perspective enables one to see that the term 'placebo effect' invites one to mistake what is accidental (the presence of an inert object or sham procedure) with what is essential (the context and meaning of the treatment being simulated). Moerman and Jonas are again instructive when they say that:

> 'Interesting ideas... are impossible to entertain when we discuss placebos; they spring readily to mind when we talk about meaning[609]'.

This shift in perspective, to meaning-theories of placebo, can perhaps best be appreciated through a number of examples:

Consider a patient who believes that their condition is the result of 'moral error, sin; demonic possession'; it is surprising to learn that 'prayer, restitution; demonic exorcism' can help the patient's condition[610]. Conversely, it is surprising to learn that commonly prescribed treatments are also effective for seemingly superfluous reasons. For example, antacid tablets help ulcer disease, but more tablets (not a different dosage, simply a different number) have a greater therapeutic effect[611]. The surprise comes from wondering how it is that prayer, homeopathic remedies or the mere number of tablets given can have an effect. All these examples appear much less surprising once it is appreciated that the meaning of a treatment can play a genuine therapeutic role: that one can respond to the ritual aspect of prayer and the medico-cultural association that 'more pills are better'.

[607] (D. D. Price et al., 2008) p. 567
[608] (Moerman, 2002b)
[609] (Moerman & W. B. Jonas, 2002) p. 457
[610] This example is from: (Kirmayer, 2004) p. 35
[611] (Moerman, 1983, 2000)

Consider a further example: Arthur and Elaine Shapiro comment that practices like bloodletting and 'leaching' possessed 'a common underlying theme… [which was] the removal of the bad, evil, or diseased, both physiological and psychological; this rationale reassured patients mobilizing their hope and helping them feel better[612]'. It is the fact a rationale such as this exists, rather than the explanatory validity of that rationale, that is important in understanding any effectiveness the procedure may have[613].

Reconceptualising placebo effects as being fundamentally about *meaning* offers a more coherent and analytically useful way to think about these kinds of effects. As a result this has prompted a variety of new terms to be invented for what previously have been called placebo effects. These new terms share the common idea that it is the form, rather than the content, of an intervention that generates placebo effects. For instance, Franklin Millar and Ted Kaptchuk conceptualise placebo effects under the term 'contextual healing'[614]; Pekka Louhiala and Rika Puustinen advocate the term 'care effect'[615]; and Daniel Moerman rejects the term 'placebo effects' in favour of the 'meaning response'[616]. In each case 'placebo' is replaced with context, care or meaning; thereby shifting focus away from inert objects or sham procedures.

These more expansive terms however have lead to the criticism that they are 'conceptually sloppy and heuristically befuddling[617]' and that too many disparate elements are being unhelpfully amalgamated[618]. One pertinent question then is whether there is any particular phenomenon worth naming at all; whether that name is placebo effect, contextual healing, or care effect etc[619]. It is this question that will be addressed in Chapter 10.

9.4 Summary

The research literature about placebos and placebo effects has been reviewed. I have argued that it speaks against some intuitive views one might hold

[612] (A. K. Shapiro & E. Shapiro, 1997a) p. 228
[613] The idea of therapeutic rationale will be discussed further in §11.2
[614] (Kaptchuk, 2002; F. G. Miller & Kaptchuk, 2008)
[615] (Louhiala & Puustinen, 2008)
[616] (Moerman, 2002b; Moerman & W. B. Jonas, 2002)
[617] (Lichtenberg et al., 2004) p. 551
[618] (J. J. Thompson et al., 2009)
[619] (Nunn, 2009a, 2009b, 2009c)

about placebo effects; namely that they are merely psychological phenomena, that they point to problems in the biomedical paradigm, and that placebos are inert substances responsible for generating placebo effects. Instead it has been shown that a wide range of factors are responsible for generating placebo effects, and they can do so through specific physiological mechanisms. It has also been shown that, in general, the magnitude of placebo effects is highly variable; they can be most convincingly detected in more subjective conditions, such as pain. The extent to which treatment components interact with other components of a treatment is not clear however. At the very least evidence suggests that the Additivity assumption does not hold generally, and therefore cannot simply be assumed in any particular case. Finally it has been argued that the most coherent attempt to understand placebo effects involves understanding them in terms of the meaning, symbolism and significance that treatments and the treatment context has for patients.

The resulting picture of placebos and placebo effects is quite different. 'placebos' seem hardly to matter at all; there is no single 'placebo effect'; but there are multiple mechanisms by which such effects are generated.

CHAPTER 10

10. Placebos and the logic of placebo comparison

Chapter 9 has set out a modern but conventional view about how to conceptualise placebos and placebo effects. First some prior intuitions that one might have about placebos and placebo effects were dismissed. Second a positive account of placebo effects from the contemporary literature was described, in which placebo effects are conceptualised in terms of physiological responses to the meaning of a treatment or therapeutic context. Such an account draws on a large and growing body of both experimental and theoretical research. Even so it is also conventional in the sense that the concepts of 'placebos' and 'placebo effects', while acknowledged as potentially problematic, are not fundamentally questioned. The research literature presupposes that it is at least acceptable to talk about placebos and placebo effects, even if those terms are a little fraught[620]. In what follows however I argue for a stronger view about how to think about placebos and placebo effects. I argue that the terms placebo and placebo effect should be abandoned.

Robin Nunn has also recently argued that the terms 'placebo' and 'placebo effect' should be abandoned: he hopes for a post-placebo paradigm in medicine[621]. He claims the terms are confused, and that there is good empirical evidence that lumping a disparate range of elements together under these terms only adds to the confusion. The point being that, if one wishes to say something informative about medical treatments, 'placebo' and 'placebo effect' are not terms which are analytically useful. Instead, one should always be much more specific about the particular details of particular therapeutic situations; which as a result removes the need to use the terms 'placebo' or 'placebo effect'.

I agree. In what follows I argue in support of Nunn's position. Nunn argues that abandoning these concepts is both possible and preferable: I think that much of the work needed to support Nunn's position can be achieved by considering the logic that underlies placebo comparisons. My argument is about the use of the term 'placebo' in a research context. If the term is valid anywhere, one might expect it to be valid in the context of a placebo controlled trial. However I expect the argument

[620] See for example the recent placebo theme issue of *Philosophical Transactions of the Royal Society B, Biological Sciences*. 2011:336(1572).
[621] (Nunn, 2009a, 2009b, 2009c)

to apply to the clinical context also, since the general idea is simply that the term obscures what should be explained in more precise terms. In Chapter 11 some consequences for the way placebos are understood in a clinical context will be sketched.

The argument below goes as follows: Like all comparisons, placebo comparison is just a case of comparing one thing with another, but it is a mistake to think of placebo comparison as a case where something is compared to 'a placebo'. Placebo comparison should be understood as a situation which sets-up the experimental groups in a particular way; not as a situation involving objects or procedures called 'placebos' employed in order to control for 'placebo effects'.

In essence my argument is an elaboration of a simple idea, which is neatly summed up by Austin Bradford Hill:

> 'To some patients a specific drug is given, to others it is not. The progress and prognosis of these patients are then compared. But in making this comparison in relation to the treatment the fundamental assumption is made – and must be made – that the two groups are equivalent in all respects, except for the difference in treatment[622]'

10.1 Placebo comparison

I claim that the key *epistemic* aim of placebo comparison, which is what is important to this discussion, is to learn about the efficacy of particular aspects of a treatment. That is not to say that there might not be other aims in mind when placebo comparisons are performed; such as having to meet regulatory requirements on the road to getting a new treatment approved, or performing a trial that is more likely to show a new treatment in a positive light (as opposed, that is, to comparing with the current best treatment). These other, more instrumental, aims will not be the focus of my argument however.

It should be noted that while I claim the aim of placebo comparison is to investigate efficacy, placebo comparison is not, by any means, the only way to learn about efficacy. That, to reiterate, follows from the conclusions of Part Two: good

[622] (Hill, 1951) p. 278

evidence does not *only* come from controlled clinical research. However placebo comparison is, prima facie, a good way to investigate efficacy. A principle such as 'a treatment is efficacious if and only if it outperforms placebo' looks very tempting: it underlies, for example, the often rehearsed argument that PCTs possess unparalleled 'assay sensitivity'[623], though whether placebo comparisons really do possess significant epistemic virtue, over and above other comparisons, has recently been questioned[624]. In the ideal case at least, the logic of placebo comparison is well-equipped to give us insight into the efficacy of a treatment.

In this first section I describe the logic of placebo comparison and the role that 'placebos' are supposed to play in it. I then argue that what counts as a placebo group depends entirely on the comparison being performed: 'placebos' are not logically prior to particular comparisons. In the second section, §10.2, I put forward a view of placebo comparison that removes reference to 'placebos'.

10.1.1 The logic of placebo comparison

The paradigm case of placebo comparison is the PCT of a drug. Such a comparison is done in order to measure the capacity of the drug contained in the treatment to produce therapeutic effects. To avoid confusion and to make clear what is meant by talking in terms of 'components of a treatment' we can stipulate a distinction between drug and treatment. Take 'drug' to denote the (allegedly) therapeutic chemical or chemicals, and take 'treatment' to denote a delivery system, perhaps but not necessarily containing a drug. Hence for clarity I mean to set-up the terms such that drugs are not pills, but treatments can be pills (though of course things besides pills can be treatments), and a pill may or may not contain a drug while still remaining a treatment, etc. The drug is merely one component of the treatment. The definition of treatment can also be widened to include not only the object which is delivering the drug, but also the way in which it is delivered. So for instance the kindness of the healthcare professional, or the patient's feeling of hope, can be thought of as some of the *contextual components* of a treatment, just as a drug is a *pharmacological component* of a treatment. Consequently: *the efficacy of*

[623] (Temple & Ellenberg, 2000)
[624] (Howick, 2009)

the drug (for condition X) is the *component of the treatment* that we wish to investigate in a *PCT of the drug* (for condition X).

The logic behind placebo comparison is straightforward, especially when put in terms of trials of drugs. In the ideal case two groups are compared which are identical in all therapeutically relevant respects, but for the fact that one group receives the drug whereas the other group does not. This is precisely the point expressed in the quotation from Austin Bradford Hill, above[625].

Note that the presence and absence of one component (in this example: a drug) is what is being compared. The comparison between the presence and absence of a treatment is a different comparison (this point will be returned to in Chapter 11). The point of comparing two groups that differ only in regards to the presence of a drug is that it allows one to infer that any differential effects between the groups can be attributed to the drug's action. This therefore allows one to reasonably claim that the drug caused those differential effects. Indeed in the ideal case this method is, as Nancy Cartwright calls it, a 'clincher', meaning that the causal conclusion is deductively implied[626]. Of course, outside of the ideal case one never compares groups that are identical in all but one respect; the best one can do is try to eliminate differences that are likely to have some unwanted confounding effect. From the point of view of the logic of the comparison however, the practical problem is irrelevant.

This logic can be generalised beyond trials just of drugs. The efficacy of a drug is only one component of a treatment, and there are many different components that one might wish to investigate the efficacy of, beside a treatment's drug content. The logic of placebo comparison is indifferent to whether the particular component to be singled-out happens to be a treatment's drug content. For example, consider the following case where one investigates whether a treatment consisting of a pill containing x mg of drug performs better than a treatment consisting of two pills, one of which contains x mg of drug and the other of which is a sugar pill. In that case, it would be the efficacy of 'receiving an extra sugar pill' that would be the component of the treatment being investigated; because that is the component of

[625] Or in other words (where 'C' is the cause, and 'E' the effect): 'Roughly, an RCT is ideal iff [that is, if and only if] all factors that can produce or eliminate a probabilistic dependence between C and E are the same in both wings except for C, which each subject in the treatment group is given and no-one in the control wing is given, and except for factors that C produces in the course of producing E, whose distribution differs between the two groups only due to the action of C in the treatment wing.' (Cartwright, 2009) p. 64

[626] (Cartwright, 2007, 2011a)

the treatment that has been singled-out. Notice also that this is a contextual, not a pharmacological, component of the treatment.

The logic of placebo comparison simply involves singling out particular components of treatments, to which one may or may not be able to attribute efficacy. There is no logical requirement to only attribute efficacy to the action of drugs. I claim that this should be uncontroversial: it really is nothing more than an elaboration of the Hill quote above.

10.1.2 Where do 'placebos' enter into the logic of placebo comparison?

Consider again the influential but much criticised[627] definition of a placebo put forward by Arthur and Elaine Shapiro (quoted above in §9.1.3)[628]:

> '[A placebo] is any therapy prescribed knowingly or unknowingly by a healer, or used by laymen, for its therapeutic effect on a symptom or disease, but which actually is ineffective or not specifically effective for the symptom or disorder being treated[629]'

Consider that this definition entails that 'patting ones head' might be a placebo for pain relief, if a clinician recommended it as a supposedly effective treatment. The fact that the clinician recommends it as a treatment for pain reliefs fulfils the first part of the definition. The second part is fulfilled because, as is intuitively clear, 'patting one's head' is (at least under usual circumstances) actually ineffective for treating pain relief: more likely it will make it worse.

'Patting one's head' however would be entirely useless in a PCT of aspirin, despite the fact that, according to the Shapiros' definition, it is a placebo pain relief treatment. The reason it would be useless is clear from above. To reiterate explicitly, comparing 'patting one's head' with aspirin is not a comparison which singles out only the effect of the particular aspect of the treatment that is under investigation: namely, the action of the drug aspirin. Even if one takes the Shapiros' definition seriously, the fact that something might, according to that definition, be a 'placebo'

[627] (A Grünbaum, 1981; A Grünbaum et al., 1991; Gøtzsche, 1994, 1995; F. G. Miller & Kaptchuk, 2008; Moerman, 2002b)
[628] See different versions of it in: (A. K. Shapiro, 1964, 1968; A. K. Shapiro & E. Shapiro, 1997b)
[629] (A. K. Shapiro & E. Shapiro, 1997a) p. 12

treatment for condition X, does not guarantee that it would be useful in a PCT of some other treatment for condition X.

The reason it is instructive to look at the Shapiros' definition of a placebo – even though it is highly criticised – is that it embodies an intuitive idea about 'placebos'. Namely, the idea that 'placebos' are particular things, or in other words, that it makes sense to claim that such-and-such is 'a placebo'. Such an idea is by no means unique to the Shapiros'. The head-patting example serves as a counter-example to that idea more generally. The assumption underlying any definition of 'a placebo' is that 'placebos' are conceptually prior to placebo comparisons: as if it were possible to take a jar of 'placebos' off the shelf, ready to use in some forthcoming PCT. I claim that this is false. For any candidate definition of 'a placebo' it is possible (a) to find an object that would fill the definition, but (b) compare it with another treatment for the same condition and therefore (c) fail to produce a comparison, which follows the logic set out above, and which I claim is the logic of *placebo comparison*.

It might be argued that the head-patting example only shows that there are such things as bad placebo comparisons; so that the example is, contrary to my suggestion, an example of a placebo comparison (because it involves 'a placebo'), but a bad one (because it doesn't follow the logic). Instead I claim that we should not understand placebo comparison, good or bad, as involving the comparison of one thing, called a 'placebo', with another, called the 'active treatment'. I will argue below that whether one is performing a placebo comparison depends only on whether one follows the logic set out above, and in no way depends on whether the comparison involves particular objects or procedures that some may call 'placebos'.

10.1.3 What counts as the 'placebo group' depends on the intended comparison

Branthwaite and Cooper[630] investigated the therapeutic effect of branded packaging. They made a four way comparison of branded and unbranded, aspirin and sugar pills. They found that branded packaging consistently provided more relief from headaches: 'Branding appeared to supplement both the inert placebo and the active ingredients to produce more relief than either placebo or active ingredients alone[631]'.

[630] (Branthwaite & Cooper, 1981)
[631] (Branthwaite & Cooper, 1981) p. 1578

Their result however is not the focus here, rather it is the fact that in their paper Branthwaite and Cooper call the branded and unbranded sugar pills 'placebos' and the groups which received these sugar pills the 'placebo groups'.

This is a reasonable way to label the groups if one holds the view that 'placebos' are things, since such labelling follows straightforwardly from the 'sugar pill = a placebo' idea: sugar pills are placebos, groups given sugar pills are, therefore, placebo groups.

I claim that Branthwaite and Cooper have labelled their groups incorrectly. More precisely I claim that which of their groups one chooses to call the placebo group is, without further specification, undetermined. The reason is that, as suggested above, the placebo group identifies a group playing a particular logical role in a comparison; namely, keeping all but one of the therapeutically relevant aspects of the treatment identical. From Branthwaite and Cooper's four groups one can make a number of different comparisons, and it is only with specific reference to some particular comparison that it makes sense to invoke the term placebo group.

So: If one is interested in the differential effects due to branding, between the two groups receiving aspirin containing pills, then the placebo group in that case would be the group receiving the non-branded aspirin pills. If one is interested in the differential effects due to branding, between the two groups receiving sugar pills, then the placebo group would be the group receiving the non-branded sugar pill. If one is interested in the differential effects due to aspirin, between the two branded groups, the placebo group in that case would be the group receiving the branded sugar pill. Lastly, if one is interested in the differential effects due to aspirin, between the two unbranded groups, the placebo group would be the group receiving the unbranded sugar pill.

Equally it would make no sense, for example, to call the group receiving the unbranded sugar pill a placebo group when compared to the branded aspirin group. In that case, more than one component of the treatment is picked out, and to reiterate, the logic behind placebo comparison is to single-out only one particular component of a treatment[632]. My criticism of Branthwaite and Cooper's labelling of their groups is simply that one can pick a number of different pairs (four pairs, in fact) of their four groups which are identical in all but one respect (as enumerated above). Consequently, any particular group may or may not be labelled a 'placebo group'

[632] This idea will be modified slightly in §11.1.2

depending on which comparison one intends to make. The general point that this enumeration labours is that one shouldn't call any group a placebo group, independently of some particular comparison.

This is certainly not a re-labelling a medical researcher would likely endorse, and there is a clear objection to consider here. It is an objection to the claim that as long as the two groups being compared are identical in all but one respect, then it is a placebo comparison - Isn't it just wrong to claim this? Won't any sensible researcher object that a comparison, say, of branded versus unbranded aspirin is no more a placebo comparison than a comparison between 5mg and 10mg of a drug: these are more properly called 'active' comparisons. If two groups were to receive aspirin-containing pills, and those groups differ only in respect of whether or not the pills were branded, then neither group has received a placebo. Therefore, it is not a placebo comparison; despite what I might choose call the underlying logic of that comparison.

Such an objection would seem to rest on the known 'activity' of aspirin, namely the fact that aspirin pills contain a chemical (2-acetoxybenzoic acid) with a well understood analgesic effect, whereas placebos are not thought of as containing pharmacologically relevant chemicals. The objection seems to rely on the idea that one can distinguish placebo from non-placebo components of a treatment by their mechanism. The active pharmacological components, like aspirin-content, work through a known chemical and biological mechanism, and, so the argument goes, the placebo components work through placebo-mechanisms that are relevantly different enough to justify making a distinction between active and placebo comparisons. More needs to be said about 'placebo mechanisms' for the objection to be convincing. In fact, research into – so called – 'placebo effects' provides some apparent support for this.

Recall from Chapter 9 that recent placebo research has conceptualised 'placebo effects' in terms of 'the psychosocial context surrounding the patient and the effect that this context has on the patient's experience, brain, and body[633]'. This idea recognises that it is the meaning or the symbolism which treatments have for patients that is important for generating 'placebo effects', and as was shown in

[633] (D. D. Price et al., 2008) p. 567

Chapter 9, meaning-theories are the most coherent way to understand the placebo literature[634].

The point to note here is that one might therefore argue that placebo-mechanisms have in common the fact that they involve a response to the meaning of some component(s) of a treatment; which, as the objector to my claims will argue, is a fact that provides sufficient basis to distinguish between responses generated in those ways, and responses generated by, for example, pharmacological components. Placebo comparisons, as anyone pressing the objection would reiterate, are those comparisons where the observed effects in one or both groups are generated in response to the meaning of the treatment. Sugar pills are called placebos, because the only conceivable way they could have a therapeutic effect is through these meaning-based placebo mechanisms. A comparison between 10mg and 15mg of a drug is just not that kind of comparison, and at most, a comparison of branded and unbranded aspirin-containing pills could be thought of as involving a 'pharmacologically enhanced' placebo[635].

I claim this view is not tenable. If placebo comparisons are characterised by the presence of objects or procedures which are generating their therapeutic effects in virtue of the meaning attached to them, then (as illustrated in the head patting example above) the simple fact that one is performing a placebo comparison, in that sense, need not entail that one is following the logic set out above.

Instead the objector now needs to ask of any given placebo comparison whether it does follow the logic set out above. Hence on the objector's view, placebo comparison and efficacy testing need have no connection to each other. Rather, the objector's view is that 'placebos' are just another category of objects or procedures that may be called upon as a control in a clinical trial, the purpose of which may or may not be to investigate the efficacy of some component of a treatment. The key question to ask is what work the distinction between placebo and non-placebo comparisons is supposed to do here, given that it is not related to measuring efficacy.

I suggest that this division of comparisons into placebo and non-placebo is an arbitrary division to make. It is certainly not made on the basis that comparison with placebos allows us to attribute efficacy to components of a treatment, whereas

[634] Most notably: (Moerman, 2002b) See for further examples: (Moerman & W. B. Jonas, 2002; Moerman et al., 1979; J. J. Thompson et al., 2009) (Kaptchuk, 2011)
[635] The notion of an 'enhanced' placebo occurs, for example, in: (Kaptchuk, 2002)(Kaptchuk et al., 2006)

comparison with 'active' treatments does not. Because as set out above, on the objector's view, whether a comparison is with 'a placebo' or not has nothing to do with whether the aim of that comparison to measure efficacy.

One could, equally well, stipulate to divide comparisons into those involving treatments with a component that works through the renin-angiotensin system (e.g. the ACE inhibitors – drugs such as ramipril etc). That distinction too has nothing to do with efficacy testing, and it too divides comparisons according to the mechanism by which therapeutic responses are generated. The point is that it serves no useful analytical purpose to divide clinical trials into those featuring controls that work through the renin-angiotensin system and those that do not, based on the presence or absence of, for example, ACE inhibitors in the trial. Just as, I claim, it serves no useful analytical purpose to call a highly heterogeneous set of objects or procedures 'placebos', and to then divide clinical trials into placebo and non-placebo controlled, based on the presence or absence of such objects in the trial. Given the diversity of biopsychosocial factors and physiological mechanisms that 'placebo effects' are supposed to encompass, note further a division based on 'meaning-mechanisms' is also rather less simple than a division based on the renin-angiotensin system.

Contrary to this I suggest that, if one takes the *logic* of placebo comparison seriously, then one doesn't need to talk about 'placebos' at all.

10.2 Placebo comparison without 'placebos'

10.2.1 Being specific about the details of the placebo group

Placebo comparisons are those which compare two groups that are identical in all but one respect. How this identity is ensured, or approximated to, is a question of trial design. The placebo group in a PCT needs to be designed so as to ensure the required identity, and as illustrated above, a group which is told to 'pat one's head' is certainly not a legitimate placebo group for a PCT of the drug aspirin. Fairly obviously however, a group given sugar pills exactly like the aspirin-containing pills has much more potential to be a legitimate placebo group in a PCT of aspirin. The question of what objects or procedures are necessary for any particular PCT depends on the nature of the treatment as a whole, and the component of that treatment which is being investigated.

The common equation of ‘placebos’ with sugar pills is readily explainable by the fact that pills are a paradigmatic example of a drug delivery system. It is almost too obvious to state that if a treatment includes a pill containing a drug, then it makes sense to give patients in the placebo group an exactly similar non-drug-containing pill in order to avoid confounding the therapeutic action of the drug with the therapeutic action of simply giving a pill. The fact that this is so obvious makes it possible to underrate its significance. It tempts one to make the mistake of trying to identify placebos with sugar pills, rather than taking the correct view that, across many circumstances, sugar pills are merely highly apt to ensure identity between the treatment and placebo groups with respect to ‘receiving a pill’. Sometimes sugar pills are given to a placebo group in order to meet the requirement that treatment groups should be identical in all but one therapeutically relevant respect: they are not given because they are ‘placebos’.

To reiterate: sugar pills are not a special kind of object called ‘placebos’: but sugar pills are a particularly easy to grasp example of an object that might do the work of controlling for certain therapeutically relevant aspects of a treatment in a PCT of a drug. There is no such thing as ‘a placebo’, but there are certain ‘control roles’ that need be played in placebo comparisons, just as in any meaningful comparison. If placebo comparisons are a special kind of comparison, it is not because they involve comparison with a special kind of object (‘a placebo’), but because they involve a control group (the placebo group) with special features. Those special features have been explained already: they are those that ensure the placebo group is identical to the treatment group in all but one respect.

If one wishes to investigate the efficacy of an extra 5mg of drug, on top of 10mg, one can perform a placebo comparison that compares two groups identical but for receiving either 10mg or 15mg of a drug. It is a question of how the placebo group is set-up that matters, not what particular objects or procedures are employed. Sometimes placebo comparison may involve a placebo group which receives a pill containing 10mg of a drug as a control, because the efficacy of a marginal 5mg above this is what is being investigated. At other times placebo comparison may involve a placebo group which receives a pill containing only sugar as a control, because the efficacy of a drug above the efficacy of pill-receiving is what is being investigated. Both warrant being called placebo comparisons. There is no

distinction worth making between the two that would make one a placebo comparison, and the other not.

It might be argued that 'placebo' is simply a shorthand way of labelling an experimental control such as an 'exactly similar non-drug-containing pill'[636]. This would be an argument for the view that the notion of 'a placebo' does in fact make sense, when restricted to the context of some particular comparison and on the understanding that 'a placebo' in one context may not remain 'a placebo' in another. Or put another way, one could stipulate that, in circumstances, C, object X is a placebo. Such a view asserts that the term 'placebo' is not meaningless or unhelpful. On the contrary it purports to do the helpful work of summing up important details about the control being used in a given comparison; and neither does it involve distinguishing objects and procedures on any mechanistic basis. It is merely a shorthand stipulation.

This 'placebo-shorthand' view does not succeed. A placebo group is a group which possesses the specific features which ensure identity to the treatment group, except with respect to the component under investigation. Now ask, what is the term 'placebo' supposed to go shorthand for? – Presumably, it should go shorthand for some set of measures that have been taken to ensure the identity between groups, but what set? – If it is the set of measures taken to ensure the identity of *all* the therapeutically relevant components of the treatment, besides the one being investigated, then that already has a name, it is just the placebo group. Of some purported placebo group, the key question is whether it genuinely does possess the features that would enable a legitimate placebo comparison to be made. That consists of asking questions about particular components of the treatment, such as whether the delivery mechanisms are the same, whether the patients are given the same information, whether the healthcare team have the same expectations for the two groups etc, and importantly there are no questions, at this level of specificity, that involve talking about 'placebos'.

If however the term 'placebo' is stipulated as a shorthand for some proper subset of measures, then that fails to be helpful. One still needs to ask, for the measures taken, specific questions about whether particular components of the treatment are the same between groups. Moreover the knowledge that there is identity between the groups in *only some* therapeutically relevant respects, still does

[636] or perhaps in suitably different circumstances, the act of 'patting one's head'.

not guarantee the legitimacy of the placebo group as a whole. The legitimacy of the placebo group depends on *all* components (but the one under investigation) being identical between groups. So for example, if one were conducting a PCT of the drug aspirin, delivered in pill form, one could choose to call the exactly similar non-drug-containing pills 'placebo pills'. Never the less one could still fail to conduct a legitimate placebo comparison with these 'placebo pills'; perhaps because the two groups were, say, given very different information and reassurance as part of their respective treatments. Furthermore, just because they had been called placebo pills, would not remove the need to ask specifically, whether they were similarly coloured, shaped, or possessed no relevantly-active content – which are questions one must ask anyway, even if they hadn't been called placebos.

The placebo-shorthand view fails because it has no bearing on the questions that need to be asked of a placebo comparison, in order to ensure it is a good one. One could certainly stipulate to call certain kinds of control measures 'placebos' as a shorthand, but only because any number of such redundant shorthand stipulations could be made. The key point is that it is the specific details of the placebo group, as a whole, that matter for placebo comparison. The fact that one could stipulate that a certain subset of features possessed by a particular placebo group should be called 'a placebo' does not solve any problems. It is redundant to call anything 'a placebo', even with respect to some particular comparison.

To ensure a legitimate placebo comparison has been performed one must ask questions about all the therapeutically relevant components of a treatment. As seen in Chapter 9, the components which turn out to be relevant can be unintuitive. This included: the mere number of pills[637], the branding of pills[638], whether one is given a pill or an injection[639], the justified belief that one has undergone surgery[640], verbal suggestions and the attitude, enthusiasm and behaviour of the healthcare team[641], and the cognitive and emotional states of the patient[642]. The properties that some set of objects or procedures will need to possess to ensure that some comparison is a genuine placebo comparison will depend entirely on the details of

[637] (Blackwell et al., 1972; A. J. M. de Craen et al., 1999; Moerman, 2000)
[638] (Branthwaite & Cooper, 1981)
[639] (Amanzio et al., 2001)
[640] (Cobb et al., 1959; Dimond et al., 1960)
[641] (H. M. Adler & Hammett, 1973; Blasi et al., 2001; Kaptchuk, 2002; Ong et al., 1995; D. D. Price et al., 2008)
[642] (Stewart-williams & Podd, 2004)(D. D. Price et al., 2008)

the component of the treatment being investigated. For the reason that, how the identity between groups is achieved will obviously differ according to the nature of the treatment and the component of interest. There is a danger associated with calling certain objects or procedures 'placebos', in so far as this tempts one to forget to check they are genuinely ensuring the required identity between groups[643].

In spite of this a medical researcher may still object that the terms 'placebo' and 'placebo effect' are perfectly functional, even if they are problematic. They may argue that unless the terms are leading to clinically meaningful mistakes being made, then my argument, in an important sense, does not matter. In response I would suggest that the terms may well introduce practical problems. Talk of 'placebo effects' carries with it the implication that the effects are unreal, or in some way mysterious (the quotes at the beginning of Chapter 9 illustrate this well). Clinicians are likely to be able to do more to help their patients if 'placebo effects' are not a black box.

More substantially, talk of 'placebos' can tempt one to neglect questions about the adequacy of the placebo group to ensure the required identity to the treatment group[644]. To give one example: the credibility of trial results are often diminished where blinding has been unsuccessful[645]. If identical-looking pills given to both groups differ, say, in taste or side-effects, then this introduces a reason to worry about the success of the trial being blind. I admit that the extent to which this is a clinically meaningful problem is an empirical question; never the less, being explicit about how the control group was set-up is a matter of rigour. Talk of 'placebos' obscures legitimate questions about the specific details of the control group.

I have argued that it is the logic of placebo comparison that dictates the nature of the controls to be used when one sets out to measure efficacy. The implications of this are less readily acknowledged: there is no sense besides arbitrary stipulation in calling an object or procedure, which in certain circumstances can do some of that work, a 'placebo'. Once one knows that a placebo comparison is being performed, there is no further need to invoke the term 'placebo'. Rather, the

[643] The same point was also noted in Chapter 9. See (Golomb, 1995; Golomb et al., 2010)
[644] Use of the term 'placebo' could also be important in a different way, if that usage created therapeutically relevant expectations in a patient. For example in a clinical context, through being told one is receiving 'a placebo'; or in a research context, through being enrolled in a trial and told that one may be randomised to a placebo group. See for example: (Enck et al., 2011; Kaptchuk et al., 2010)
[645] (Rabkin et al., 1986)

meaningful questions to ask involve being specific about the details of the controls – so that one can evaluate the plausibility of alternative explanations of the results.

10.2.2 Is the failure of Additivity in general problematic?

Chapter 9 also included a discussion of Additivity and noted that in general it cannot be assumed. It might be argued that the failure of Additivity poses problems for the argument developed here. Consider that a placebo comparison of 0 and 5mg of a drug may well yield an effect size different from a placebo comparison of 10 and 15mg of the same drug (perhaps due to some threshold effect). Consider further that a placebo comparison of 0 and 5mg of a drug may also yield a different effect size as some other component of the treatment varies; hence a placebo comparison of 0 and 5mg in one context may yield a different result than in another (due to some interaction effect, or overdetermination effect). If one cannot generally assume Additivity, then, so the argument goes, comparing the average effect sizes of different groups requires additional evidence to ensure one is accurately measuring the efficacy of the component in question.

This is not a strong challenge to the argument above. At most the, in general, failure of Additivity speaks against the view that placebo comparisons provide an 'absolute' measure of the efficacy of treatment component. I claim however that the account of placebo comparison given above does not imply such a view. It was never claimed that placebo comparison measures the absolute efficacy of a treatment component. Indeed, others have argued that the failure of Additivity, in general, provides a reason against holding the view that placebo comparison does measure absolute efficacy[646].

More importantly however, recognition of the fact that Additivity cannot be assumed serves to highlight the fact that extra evidence is needed in order to export the results of a given placebo comparison to other circumstances. Placebo comparisons do not provide evidence that the effect size of the treatment component is robust across other circumstances; but the view above does not imply the contrary. The fact that Additivity cannot be assumed cautions against overstating the claim that placebo comparison measures efficacy, but it does not undermine that claim. This idea will be returned to briefly below, and also in Part Four.

[646] See also (Howick, 2011) pp. 107-112

10.3 Summary

I have argued for abandoning the term ‘placebo’ in a research context. I claimed that when one considers the logic that one tries to follow when performing a placebo comparison, there is no role for ‘placebos’ to play. It is identity between the treatment groups (in all but one respect) that matters. In general terms the key point made in this chapter is that the level of specificity and rigour required to perform a legitimate placebo comparison is a level at which one does not need to use the terms ‘placebo’ or ‘placebo effects’; indeed those terms obscure assessments of a comparison’s legitimacy.

As explained at the beginning of this chapter, I think the this argument supports the position advocated by Robin Nunn, who holds the view that we should abandon the concept of ‘placebos’ and ‘placebo effects’ altogether. Perhaps the support is not total however, since I am happy to use the term ‘placebo comparison’, whereas Nunn is not. This difference is not significant. The conclusion is that (what I would like to call) placebo comparison involves no commitments to ‘placebos’ or ‘placebo effects’.

CHAPTER 11

11. Implications of the arguments about 'placebos'

The motivation for looking more closely at placebos was interest in the crucial role that placebo controls play in debates about homeopathy. The Canonical Criticism makes essential use of placebos as a special evidential standard. Placebo controls define the threshold that a treatment must exceed in order to 'work' legitimately.

Chapter 9 introduced some counter-intuitive empirical results from contemporary research into placebo phenomena. Most importantly Chapter 9 sought to show how this research literature provides empirical evidence against some common ideas about 'placebos'. The account put forward was conventional (in a sense was explained at the beginning of Chapter 10), and it stressed the reality and diversity of 'placebo effects'. In doing so, Chapter 9 provided the groundwork for the argument of Chapter 11. Chapter 10 argued that reflection on the logic of placebo comparison shows that we should abandon the terms 'placebo' and 'placebo effect'.

The purpose of this chapter is to draw out further and more fully the implications of the argument made in Chapter 10. The concern will firstly be with the implications of the argument, for the concept of efficacy and the distinction with effectiveness. Consider that if one holds the view that the effectiveness of a treatment (over and above the natural course and variation of the condition) is just the efficacy of the treatment plus any placebo effects, then one obvious question to ask is what implications the argument of Chapter 10 has for such a view. How should one re-evaluate the distinction between efficacy and effectiveness if one is to abandon reference to 'placebos' and 'placebo effects'?

Secondly, this chapter will examine the implications of the argument of Chapter 10 for views about the ethics of placebo treatments. Consider that if one holds the view that placebo comparison represents an important ethical standard that treatments must meet – for instance, the view that it is unethical to provide patients with 'placebo treatments' – then, again, what implications does the argument of Chapter 10 have for such views. What is the ethical significance of placebo comparison?

11.1 Efficacy and effectiveness

One view about the distinction between efficacy and effectiveness was implicit in the discussion of the evidential debate about homeopathy in Part One. The view is simply that a treatment is efficacious when it is better than 'placebo', and effective when it is better than no-treatment. The view implies that efficacy and effectiveness both measure the ability of a treatment to produce an effect above some level; but that level differs, because the comparison that is being made differs. Furthermore, the effectiveness of a treatment is (perhaps) necessary but not sufficient for its efficacy[647]. The STC explain the failure of the sufficiency claim when they state: 'The answer to why a medicine can be effective without being efficacious lies with a phenomenon known as the placebo effect[648]'

A second view is that the distinction between efficacy and effectiveness is a distinction between controlled and real-world circumstances. This view can also be found in the STC report[649]. On this view, a treatment is efficacious if outperforms 'placebo' in randomised trials and effective if it is useful in clinical practice. This distinction has little to do with 'placebos' or 'placebo effects', or more generally, the different comparisons one might make. On this view efficacy and effectiveness do not measure the same thing. Rather, it seems that efficacy is, as above, measuring the ability of a treatment to produce some effect (but it is open as to what comparison is being referred to), whereas effectiveness is supposed to provide some measure of that treatment's usefulness. Effectiveness is neither necessary nor sufficient for efficacy, on this view. Against sufficiency, on this view and as above, effectiveness is possible without efficacy when the treatment is a 'placebo'. Against necessity, a treatment may be efficacious but not effective for a range of reasons. For example, in the case that the treatment is efficacious but has a small absolute and relative effect-size, and produces particularly undesirable side-effects; that is, it is efficacious but clinically useless. Or alternatively for example, in the case where other factors in the real-world defeat the efficacy of the treatment's otherwise efficacious components[650]. This second example is similar to a failure of external validity: the

[647] The STC hold that effectiveness is neither necessary nor sufficient for efficacy (House of Commons Science & Technology Committee, 2010) see table in para 28 and paras 26-29 (p. 9)
[648] (House of Commons Science & Technology Committee, 2010) para 29 (p. 9)
[649] (House of Commons Science & Technology Committee, 2010) paras 26-28
[650] This is only partly related to the failure of the Additivity Assumption, in general.

treatment is efficacious, but highly sensitive to circumstances; it turns out to be useless in real-world circumstances[651].

How the argument from Chapter 10 affects these two views will be considered below. It is worth noting first however that other authors have also considered the implications of 'placebo' research for the distinction between efficacy and effectiveness. Even if one does not endorse the argument of Chapter 10 (that the terms 'placebo' and 'placebo effect' should be abandoned), some revision of one's views about 'placebos', that took insights from the empirical evidence reviewed in Chapter 9, would still be expected to have important consequences. Just such an argument is made by Harald Walach[652]. Notably, he posits a 'paradox of efficacy' which needs to be resolved, he argues, by a revision of the distinction between efficacy and effectiveness. In fact, Walach's argues that his paradox necessitates a radical revision of how efficacy is conceptualised. Before discussing the two views above, it is important and illustrative to discuss (§11.1.1) Walach's more fundamental challenge to the distinction between efficacy and effectiveness. I argue below that Walach's paradox is in fact no paradox at all, then in §11.1.2 I consider the distinction between efficacy and effectiveness more directly.

11.1.1 Walach's paradox of efficacy

Walach's claims that there is a paradox in the way that efficacy and effectiveness are typically conceptualised, because 'placebo', or non-efficacious, treatments can be more effective than 'non-placebo', efficacious, treatments. Consider two different medical treatments A and B, which treat the same condition. Walach claims that a 'paradox' arises in cases which are described by the following apparently contradictory set of true statements: (1) A is equivalent to placebo; (2) B outperforms placebo; (3) A outperforms B. This contradiction, according to Walach, permits one to generate the following paradoxical statement:

[651] As Cartwright has noted, we need independent evidence for these claims, see: (Cartwright, 2009) (Cartwright, 2011a)
[652] (Walach, 2011)(Walach, 2001)

(P) 'A non-efficacious treatment could be more effective than an efficacious one[653]'

That such a case described in (P) is possible should be clear from Chapter 9 and Chapter 10, but more concretely Walach gives an example concerning the use of acupuncture for treating pain (arising from a number conditions)[654]. The results Walach draws on show that both acupuncture and placebo (that is, 'sham' acupuncture), which are themselves difficult to distinguish in terms of their efficacy[655] are significantly more effective than conventional (NSAID) treatment[656], which is superior to placebo[657].

Walach claims that this situation represents a challenge to the coherence the concept of efficacy. On the basis of this apparent paradox, Walach makes a very strong claim about the implications that 'placebo' research has for the concept of efficacy, namely:

> 'The placebo effect points out the cracks in our conceptual edifice of efficacy and effectiveness. The paradoxes it leads us into also suggest a way forward: to not only conceptualise efficacy as net effect against placebo, but also as general effectiveness[658]'

[653] (Walach, 2011) p. 1874 – note that the quotation is Walach's, but labelling it '(P)' is my presentation of it.
[654] See for example: (Cherkin et al., 2009; Kaptchuk et al., 2006)
[655] Note that quite what ought to count as a legitimate placebo group in acupuncture trials is not obvious (Streitberger & Kleinhenz, 1998; P. White, Lewith, Hopwood, & Prescott, 2003) – what aspects of the treatment are the potentially efficacious ones that one wants to single-out? What should the placebo group be set up to control for? - Non-penetrating 'acupuncture' (Steitberger needles); Acupressure anywhere; acupressure at a specific location, acupressure and a certain kind of twisting motion (re twisting see: (N. Goldman et al., 2010)). The legitimacy of controlling for these different aspects of the treatment will depend both on methodological points concerning whether the sham treatment can be delivered while plausibly controlling for other aspects (e.g. practitioners' knowledge of the sham, patients receiving an indistinguishable experience of the treatment) and keeping the trial blind. As well as depending on details about the proposed mechanism of acupuncture (e.g. if it is mere penetration that matters according to the proposed mechanism, then proponents will claim that a trial which singles out the location of the acupressure *could never* give a positive result – the same kind of mistake as taking paracetamol with either water or orange juice, as if that were a placebo controlled trial of paracetamol).
[656] (Haake et al., 2007)
[657] This acupuncture example is also discussed in (Howick, 2011) pp. 89-94.
[658] (Walach, 2011) p. 1871

Walach's argument for this is not at all clear. He aims to show the falsity of two assumptions which together are supposed to lead from the statements (1)-(3) to (P). These are that: 'placebo controls control for background noise that is comparatively uniform[659]', and that, 'efficacy [should be seen] only in terms of differences between active and control conditions[660]'. The falsity of the first assumption is not controversial: indeed just that point was made in Chapter 9. I claim however, contrary to Walach, that the second assumption is true; at least in so far as it expresses a point made in Chapter 9, namely, that efficacy is measured in placebo comparisons. This point will be returned to below, firstly however it is important to consider the problem with the statements (1)-(3), and then how (P) ought to be interpreted.

The key to understanding why there is in fact no contradiction in (1)-(3) is to re-describe them with reference to the different components of the respective treatments that are being investigated. Once one begins talking more specifically about the efficacy of different *components* of treatments, one can see that the contradiction cannot arise. (1) and (2) refer to two different placebo comparisons: Following Chapter 9, the details of how the placebo groups have been set-up must be described. Immediately one can appreciate that the placebo group in a PCT of some component of treatment A may be quite different from the placebo group in a PCT of some component of treatment B. To reiterate a point made previously, there are no jars of 'placebos' sitting on a shelf, waiting to be used in the next PCT. In fact, Walach's 'paradox' arises precisely because the placebo group from a PCT of acupuncture is very different from the placebo group of a PCT of an NSAID. These two placebo groups are set-up in very different ways, which is apparent once one describes them in more detail.

This extra specificity in the description shows that one cannot combine (1) and (2) to produce a statement (namely: "B outperforms A") which contradicts (3). To do so commits the fallacy of equivocation on the term 'placebo'. It provides a good illustration of why the term is unhelpful: it glosses over the *obvious* differences between a placebo group in a PCT of acupuncture and in a PCT of an NSAID. If one eliminates reference to 'placebos' and is instead more precise about the particular details of the two statements (1) and (2), one cannot equivocate.

[659] (Walach, 2011) p. 1875
[660] (Walach, 2011) p. 1875

If the contradiction in (1)-(3) can't be generated, how then should (P) be interpreted? - (P) might be thought to amount to the claim that some placebo treatments are more effective than some 'real' treatments. This would at least highlight the fact that 'placebo treatments' can be effective in their own right. However the argument from Chapter 10 allows one to improve the idea further. Re-writing the apparent paradox in terms consistent with Chapter 10 therefore, we get the entirely unproblematic:

(P') A treatment, A, possessing some non-efficacious component, X, can be more effective than a treatment, B, possessing some efficacious component, Y.

(P') is sufficient to resolve the apparent paradox that (P) present one with; but it is also completely trivial. It is sufficient to resolve the paradox because it makes explicit the different comparisons are being made. In (P') efficacy is attributed to some component of treatment A on the basis of comparison with a placebo group; effectiveness is attributed to the whole treatment A, in comparison with the whole treatment, B. Walach's paradox, (P), is really only problematic if one holds the view that a treatment with an inefficacious component must also be an ineffective treatment. One can see from (P') however that the fact that one component of a treatment may be inefficacious does not necessarily count against the effectiveness of the treatment in comparison with another treatment.

(P') is trivial because it is indifferent to which components are picked out. One can pick out many non-efficacious components of any effective treatments, which thereby satisfy (P'). Consider for example, the excipients used in pills[661], it is likely that these excipients will be non-efficacious components of pill-based treatments for very many conditions[662]. It is no surprise therefore that an effective treatment, with *some* inefficacious components, can be more effective than another effective treatment.

[661] For example: diluents such as lactose or sorbitol, or antiadherents such as talc or magnesium stearate. See: (Winfield & Kennedy, 2004) p. 230

[662] That these components are genuinely inefficacious is something that must be assessed, depending on the particular condition that is being treated. It should not be automatically assumed – as noted elsewhere, see: (Golomb, 1995; Golomb et al., 2010)

One might therefore object to (P') as a way of resolving the paradox, because it does not capture the fact that the efficacy of some components of treatments are held to be more important than the efficacy of others. A less trivial way of resolving the paradox, ought to capture the fact that one likely has some particular privileged component in mind when one states the paradox. When one claims that a treatment is inefficacious, one is unlikely, for example, to intend this to be interpreted as a claim about the power of the excipients used to produce therapeutic effects.

Some refinement of (P') is necessary and, I suggest, can be achieved by adapting a term from Adolf Grünbaum[663], who talks about treatments' *characteristic* factors. These are supposed to be the components of a treatment which make it *that treatment* specifically; or in other words, the component that characterises the treatment (for the moment, one can gloss the fact that this is relative to some 'therapeutic theory'). To illustrate: paracetamol – that is, the drug acetaminophen – is the characteristic component of paracetamol treatment for pain relief. Other components such as the size, shape, colour, excipients, or contextual factors are all *non-characteristic* components of paracetamol treatment for pain relief. They, unlike acetaminophen content, could be altered or eliminated without thereby affecting whether the resulting treatment remained *paracetamol treatment for pain relief*. The characteristic component of any given treatment may not always be so easy to identify as in this case, but examples of drug treatments such as paracetamol for pain relief clearly illustrate the idea[664]. In fact, more will be said about this idea of characteristic components below; for now however note that when put in these terms, (P') therefore becomes:

(P'') A treatment, A, possessing some inefficacious characteristic component, X, can be more effective than a treatment, B, possessing some efficacious characteristic component, Y.

(P'') provides a much better interpretation of (P). Substituting in Walach's acupuncture and NSAID example from above: the characteristic component of

[663] (Adolf Grünbaum, 1986) fig 1. See also: (Howick, 2011) p. 81-2

[664] Notice that in the case of acupuncture and homeopathy it is not obvious what the characteristic components of the treatments are. Indeed, as we saw in Part One, what counts as the characteristic component of homeopathy is contested.

acupuncture is not efficacious[665], but acupuncture treatment is more effective at treating pain than treatment with NSAIDs, which do possess an efficacious characteristic component.

There is no problem here, and certainly no paradox; however, neither is this a trivial way to interpret (P), as (P') seemed to be. In contrast to (P'), the satisfaction of (P'') is likely to be an impressive medical fact. Using Walach's example again: the non-characteristic components of acupuncture confer a greater therapeutic benefit for treating pain than treatment with NSAIDs, and this is despite the fact that treatment with NSIADs utilises an efficacious anti-inflammatory drug on top of the efficacy of its own non-characteristic components.

One could also put this in a way that emphasises the insights of meaning-theories of 'placebos', namely: the drama, context and meaning that acupuncture creates for patients (that is, its non-characteristic components) are therapeutically more beneficial for treating pain than taking ibuprofen.

Walach wishes to draw the conclusion from his paradox that efficacy should be reconceptualised to capture facts about a treatment's overall effectiveness that may be missed when it is merely claimed that the treatment's characteristic component is not efficacious. I deny that any reconceptualisation is necessary. The problem is no deeper than noting that misunderstanding arises if one neglects the fact that the efficacy of a treatment's characteristic component is independent of the effectiveness of a treatment as a whole. The insights from Chapter 10 show that by avoiding ambiguities about which components of a treatment one is making claims about, and whether one is making claims about treatment components or treatments as a whole, then there is no paradox of efficacy.

11.1.2 Two views about efficacy and effectiveness

The two views introduced at the beginning of this chapter offer divergent accounts of the distinction between efficacy and effectiveness. On the one hand, it is supposed to be a distinction between two different comparisons one might make

[665] Assuming of course that a placebo group receiving sham acupuncture is in fact a legitimate placebo comparison, as defined by Chapter 9 – As already noted, people may dispute whether sham acupuncture should be part of a legitimate placebo comparison; which is another way of saying that they dispute what the characteristic component of acupuncture is.

when measuring the ability of a treatment to cause some effect; on the other, it is supposed to be a distinction between the ability of a treatment to cause an effect in some circumstances and the usefulness or robustness of that ability across other circumstances.

Walach's paradox above, in one sense, exploited an ambiguity in the term 'efficacy'; whether it refers only to outperforming placebo, or to overall treatment effectiveness. Walach argued that this ambiguity indicated problems with the very concept of efficacy; I argued it was simply imprecise exposition. The ambiguity could be removed by being explicit about the comparisons that certain claims were meant to indicate. An analogous point can be made about the two views of the distinction between efficacy and effectiveness noted above. These two views point to an ambiguity in the term 'effectiveness'. The first view holds that effectiveness is the ability of a treatment to outperform no-treatment. The second view holds that effectiveness states facts about how useful a treatment is in various clinical situations. To anticipate: both views are compatible, as long as one avoids equivocating on this ambiguity. I propose to do so by stipulation.

Consistent with the first view about the distinction between efficacy and effectiveness, the argument from Chapter 10 entails that efficacy and effectiveness, in so far as they both refer the measurement of a treatment's ability to produce effects, do indeed measure the same thing. One might be suspicious that one consequence of Chapter 10 is that efficacy and effectiveness seem like fundamentally different properties. One might be concerned that Chapter 10 implies that 'efficacy' is only a property of treatments' components, and 'effectiveness' only a property of whole treatments.

This is not the case. The distinction between treatment components and whole treatments is not deep, and attributions of effectiveness can always be re-phrased in terms of efficacy. Consider first an uncontroversial example, then second, a generalisation from it.

Consider first the PILL Collaboration study[666], which investigated the efficacy and tolerability of a 'polypill' to treat patients with increased cardiovascular risk[667]. This polypill treatment contained four different drugs[668]. Consequently there are four

[666] briefly mentioned in Chapter 9
[667] (PILL Collaborative Group, 2011)
[668] Aspirin (75mg), lisinopril (10mg), hydrochlorothiazide (12.5mg), simvastatin (20mg).

different pharmacological components of the treatment that are of interest: moreover the interest is in their combined effects. In other words, the *characteristic component of the polypill treatment is made up of four different drugs*. For this reason, the placebo group in that trial was designed to control for all but these four different components of the polypill treatment. The efficacy of the characteristic component of the polypill is another way of saying the combined efficacy of the four drugs contained in the polypill.

The example shows that one can pick 'components' however one likes. From the point of view of the *logic* of a placebo comparison it does not matter which component is singled-out, or whether the component which is singled-out can be broken down into further components – components are not 'atomic' in that sense[669].

Second, this same idea can be pushed further. Consider that one might take an interest in the combined efficacy of *all* the components of a treatment[670]. In that case, a placebo comparison of that component would be identical to a comparison between the treatment and a no-treatment group. Measuring the combined efficacy of all the components of a treatment is equivalent to measuring the effectiveness of that treatment. In this way attributions of effectiveness to whole treatments can always be re-phrased as attributions of efficacy. When one asks 'is this treatment effective?' one could equally well ask 'is the treatment efficacious?'. The problem of course is that when one asks whether a treatment is efficacious one is not typically

[669] The components of a treatment have more in common with a resolved vector, than a dismantled Lego house. That is to say, one chooses how to resolve a vector, one doesn't choose how to dismantle a Lego house. For example, consider a simple 'inclined plane' problem in mechanics. One resolves the forces (which are vectors, of course) into horizontal and vertical components, or components that are perpendicular and parallel to the plane, depending on which is most useful for solving the problem (introducing or eliminating coefficients that are functions of the angle of inclination).

Note that if one wants attributions of efficacy to particular components to have some deeper metaphysical significance, then some theory is needed of how the effects are being generated. If one learns through a placebo comparison that a sugar-pill in circumstances C is efficacious, then in order to turn this into useful, exportable, knowledge some account is likely to be needed for why and how the effect is generated. Presumably such an account will not refer to sugar-pills per se, but to individual's expectations and cultural associations. Whilst thoroughly interesting, issues of this sort have been put aside here and in the rest of this thesis. The key point which this chapter makes is that the logic of placebo comparison allows one to attribute efficacy to any component one chooses. That point is independent of views about which are the 'right' components to break a treatment down into.

[670] This would be a slightly odd way to talk about components; but the idea is no different from the idea that every set is a subset of itself. Indeed, one could talk about 'proper' components as one talks about proper subsets.

asking whether it is better than no-treatment; rather one is in fact asking whether the characteristic component is efficacious. However, I claim, that is a problem of imprecision on the part of the questioner. To reiterate: when one asks questions about efficacy, one needs to indicate precisely what component one is interested in.

Given that asking about the efficacy of a whole treatment – and meaning by it, how the treatment compares to no-treatment – is perhaps easily misunderstood as a question about the efficacy of the characteristic component, it may therefore be useful to use the term effectiveness to indicate this. In what follows I will use the term effectiveness for such a comparison, and restrict the term efficacy to *proper subsets* of treatments' components[671]. This amounts to the distinction between efficacy and effectiveness, described in the first view put forward above.

The second view about the distinction between efficacy and effectiveness is quite different. It is however, consistent with the points noted above. Effectiveness, on the second view, refers to facts about the clinical usefulness or robustness of a treatment's effects. Efficacy, on the second view, refers to the ability of a treatment to produce therapeutic effects (that is to say, it refers to either or both of what were termed efficacy and effectiveness according to the first view). I have argued above that effectiveness, according to the first view, can always be re-phrased in terms of efficacy. Thus the first and second views are compatible with each other; it is merely unfortunate that they both use the same term 'effectiveness' to mean two different things. To make this clearer:

Some component of a treatment may be efficacious; meaning that it performs favourably in a legitimate placebo comparison. In the special case where the component in question refers to the whole treatment, the whole treatment may be efficacious; or as it is helpful to say, effective (since a placebo comparison of a whole treatment requires the placebo group to be set-up as a no-treatment group;

[671] Note that one might take this as an argument for abandoning one or other of the terms efficacy or effectiveness. Unlike the argument for abandoning 'placebos' and 'placebo effects' however, nothing substantive turns on these terms. Rephrasing descriptions to eliminate reference to 'placebos' or 'placebo effects' is non-trivial; it is not merely a case of inserting a new synonym like 'placebo response' or 'meaning response'. Rephrasing descriptions to eliminate either efficacy or effectiveness is trivial in this sense, however. The use of one or the other is a matter of stipulation. Consequently I stipulate to use effectiveness for whole treatments and efficacy for proper subsets of treatments' components.

that is, keep groups identical, except for the presence of the treatment[672]). The efficacy of some component may be clinically useful, or may be sensitive to changed circumstances. In essence there are three things one is asking: with reference to some fixed set of circumstances, firstly does some component of a treatment cause an effect in its own right? And secondly is the treatment overall better than no treatment. Also thirdly, does the effect observed in this set of circumstances export to other circumstances[673]?

The view put forward in Chapter 10 does not threaten the distinction between efficacy and effectiveness (and Walach's attempt to re-draw it does not succeed either) but it does help to draw it more precisely. Importantly §11.1 has clarified some ambiguities in the way the distinction is made. These clarifications, as well as the introduction of the notion of a treatment's characteristic component, have consequences for the argument developed below, and in Part Four.

In §11.2 I consider how the arguments from Part Three affect one's view about the ethical significance of placebo comparison. The notion of a treatment's characteristic component will be particularly useful in this regard.

11.2 The ethical significance of placebo comparison

11.2.1 The value-leadenness of the characteristic component

One consequence of the argument in Chapter 10 is that there are multiple placebo comparisons that one could perform with some particular treatment; depending only on how many different components one might choose to single-out. Indeed it was this fact that made the satisfaction of (P') trivial. Consequently in §11.1.1 it was noted that one often has a particular component in mind when talking in an imprecise way about the 'efficacy of a treatment'. This idea was captured by the

[672] Note of course that 'knowing one is receiving a treatment' is a component of the treatment – indeed, it may well be one of the efficacious components. See: (Cobb et al., 1959; Dimond et al., 1960)

[673] The third question has not been, and will not be, dealt with here. I simply note that it requires substantial work, firstly for all the familiar reasons concerning external validity, but secondly because of the, in general, failure of Additivity and the other assumptions that must be met by ideal randomised trials. It is not at all obvious how the efficacy of different components may combine with or defeat each other as circumstances change. See: (Cartwright, 2011a, 2011b; Cartwright & Mantzavinos, 2009; Cartwright & Munro, 2010) also interestingly: (Mumford & Anjum, 2011) Ch. 2

introduction, in (P''), of the notion of a treatment's characteristic component: the characteristic components picks out the component that defines a treatment as *that* treatment – for example, the paracetamol in paracetamol treatment for pain relief.

More needs to be said about this notion of a characteristic component. Below I describe the way in which the efficacy of the characteristic component of a treatment seems to be important; in a way that is not the case with other components. That is to say, the way in which it seems to matter that the characteristic component is efficacious.

A preliminary point to note is that in the examples given, the characteristic component has been easily identifiable. Since the focus of the discussion has been placebo comparison, then the assumption has been that the treatments under discussion are amenable to placebo comparisons. One obvious question is how to proceed in cases where it may difficult or impossible to identify the characteristic component: such as might be the case with 'complex interventions', mentioned in Part One. Consider for example acupuncture. In this case, that there is some characteristic component is not so much in question; but questions do arise over what the characteristic component might be. In an unhelpfully wide sense the use of needles is characteristic, but the arguments are over where they should be placed, whether they should be twisted etc. On the other hand, the Medical Research Council's guidance on complex interventions gives the example of a stroke rehabilitation unit as another kind of complex intervention[674]. In that case, it is not at all clear whether one could identify a characteristic component: it seems doubtful that there is a necessary and sufficient set of components that makes 'a stroke unit' because the intervention itself is vague.

In what follows therefore the discussion is restricted to treatments where one would legitimately expect there to be a characteristic component. This is not so restrictive as to disqualify from the discussion treatments where it is difficult to identify the characteristic component, or difficult to design experiments to measure the efficacy of the characteristic component: so acupuncture remains a relevant example, as does homeopathic treatment. It does however disqualify interventions such as stroke rehabilitation units.

Putting this point aside, I suggest that the efficacy of the characteristic component of a treatment is important, for ethical reasons, in a way that the efficacy

[674] (Medical Research Council, 2000)

of other components is not. This can be easily illustrated by considering some straightforward examples.

Consider first that the known efficacy of the characteristic component of a treatment clearly plays an important role in supplying the rationale for the clinician to provide that treatment. To take a trivial example: a clinician gives paracetamol treatment for pain relief because she knows that the characteristic component is efficacious for treating pain. Consider a further example of an effective treatment which has an inefficacious characteristic component (if one assumes that the comparison with sham acupuncture is a legitimate placebo comparison, then acupuncture is such a treatment). In this case ethical questions arise about whether the known inefficacy of the characteristic component is a barrier to providing the treatment. Would a clinician be acting appropriately if she provided a treatment with a characteristic component that was known to be inefficacious, if she provided it on the basis of the treatments overall effectiveness? Additionally, how might the fact that the efficacy of the characteristic component of a treatment is unknown affect the kinds of reasons one can give for providing it?

I do not propose to answer these questions. The minimal point made here is simply that the efficacy of the characteristic component of a treatment is value-laden in a way that other components are not. It seems to matter, that is to say it makes some ethical difference, whether the characteristic component of a treatment is known to be efficacious or not. It is still an open question what difference it makes (or perhaps whether it really does make a difference); I only make the general observation that the characteristic component seems to be normatively different from other components of the treatment.

To further illustrate: one is warranted to form certain expectations about a treatment on the basis of knowledge of the characteristic component. Consider paracetamol treatment for pain relief again. If one receives a pill containing paracetamol to treat one's pain, then one expects that the efficacy of paracetamol will be part of the explanation of why the treatment is effective overall. One does not expect this from the other non-characteristic components of the treatment, such as the excipients used. If it turns out that some excipient is efficacious for the condition being treated (in this example, pain relief), that is a 'useful bonus' to the patient and also an issue for the design of randomised trials, but one does not demand that it *should* be the case. In contrast, the efficacy of the characteristic component is

intertwined with the rationale for providing *that* treatment, rather than another. It is far more plausible to demand that it *should* be the case that the paracetamol in paracetamol treatment for pain relief is efficacious for treating pain. Putting this another way, if the characteristic component of a treatment is inefficacious, then there is a plausible sense in which, one could claim that the treatment doesn't 'work' and may be inappropriate, even unethical, to provide to patients[675]. Again, this claim is much harder to make about some non-characteristic components of the treatment. It is simply this difference that I indicate by noting that the characteristic component of a treatment is value-laden.

This idea, that the notion of a characteristic component is value-laden, will be helpful in reinterpreting the typical arguments made about the ethics of 'placebo treatments'. To anticipate that discussion, the typical view is that giving 'placebo treatments' is thought to involve deception of patients. This deception is used to anchor bioethical arguments (about violations of autonomy, harm etc) for the view that giving 'placebo treatments' is therefore unethical. Indeed, this can be seen clearly in the 'No Placebos' argument made by opponents of homeopathy. As was

[675] It might be argued that the notion of a treatment with an inefficacious characteristic component could serve as a definition of 'a placebo' (this is close to what Grünbaum was attempting by coining the phrase (Adolf Grünbaum, 1986)). Indeed, it seems that it might be more serviceable than other definitions, such as the Shapiros' already quoted above. The reason it might be more appealing is firstly that it accommodates the fact that there is no single 'placebo effect' or 'placebo mechanism'. It does so because it has no commitment to what the non-characteristic components of a treatment are; the point is simply that if a treatment works in virtue of some of those components, then it is a 'placebo treatment'. A second reason it might be more appealing is precisely because, as claimed above, the characteristic component is value-laden. Defining placebo treatments as those with inefficacious characteristic components captures the fact that 'placebo treatments', if they are effective, are supposed to be effective for the wrong reasons. Furthermore, by making reference to characteristic and non-characteristic components the definition is not anchored to particular objects or procedures, but is also a function of the treatment rationale: a sugar-pill may be 'a placebo' if it is given as a pill full of aspirin, but not if it is given as a pill full of hope.

Recall, as noted in §9.1.2, that no definition of 'placebo' can satisfy the logical role that the placebo group must play in placebo comparison. This definition is no exception: just because a treatment has an inefficacious characteristic component does not make it useful as a control in a placebo comparison. Never the less, one could reply to this with the claim that, while there is no room for a definition of 'placebo' in a research context, it would still be a useful definition in a clinical context; because it does not matter, in that context, whether the 'placebo' could or could not be used as a control.

A better argument therefore is not to deny that one could stipulate that treatments with inefficacious characteristic components are 'placebos', but to deny that this is a stipulation worth making. Even in a clinical context, I claim that it is preferable to describe the details of treatments; firstly as a point of rigour, and secondly for the kinds of ethical reasons that will be described below.

shown in Part One, opponents of homeopathy claim that when evaluating whether homeopathic treatment works, what matters is why it is effective, not merely that it is – hence the need for placebo controlled trials to test that the characteristic component really does contribute to the effectiveness of the treatment. The argument opponents give in support of this is that it would be unethical to provide homeopathic treatments, if their effectiveness turned out to come solely from 'placebo effects'. In turn, the justification for this ethical view was the claim that the effectiveness of 'placebo treatments' relied, in an essential way, on deceiving patients. And it is this deception that is the source of the ethical problem. To reiterate, the problem for opponents of homeopathy is not that homeopathic treatments are ineffective but that, even if they are effective, they rely, unethically, on deception (because they are 'placebo treatments'). It is this line of reasoning that will be examined further below.

The task here is to unpack these issues about the ethics of providing effective treatments, but – consistent with Chapter 10 – without reference to 'placebos' or 'placebo effects'. As the quotes at the beginning of Chapter 9 aptly illustrate, it is too easy to let the term 'placebo' hide the normative work that a mere placebo comparison seems able to achieve. The question then is why does the efficacy of the characteristic component of a treatment matter? – Intuitively, it would seem to have something in common with the unethical and deceptive nature of 'placebos'. If it is the case that effective treatments with inefficacious characteristic components are likely to involve some unethical deception of patients, then that provides a reason why the efficacy of the characteristic component ought to matter. In §11.2.2 I draw a connection between the view that 'placebo treatments' are unethical, and the view introduced above that the efficacy of the characteristic component of a treatment is value-laden. In doing so, I provide a reinterpretation of traditional arguments about the ethics of 'placebo treatments'.

11.2.2 Deception and treatments with inefficacious characteristic components

It is claimed that the ethical problem with the clinical use of 'placebo treatments' is that they necessarily involve deceiving patients. For example:

> ‘if there is an ethical problem in therapeutic use of placebos, the problem is that of deception[676]’

> ‘placebos given in the context of medical treatment are essentially deceptive. If deception has no place in clinical medicine, placebos have no place[677]’

> ‘the biggest barrier to the use of placebos in clinical practice is the almost universal perception that for a placebo to be effective it must be administered deceptively[678]’

This view about the ethics of ‘placebo treatments’ conceives of their use in terms of a patient being told that they are receiving an effective treatment, when in fact – if the treatment is effective at all – it is only effective because of the patients false beliefs about it. This view presents a standard picture of ‘placebo treatment’ and its problems, as follows: first, as already noted, patients’ false beliefs about the treatment they are receiving are responsible for any subsequent effectiveness of the treatment[679]. For example, a clinician may tell their patient that a saline injection “is morphine” or merely “a powerful painkiller”. Second, since those false beliefs are a consequence of some deliberate lie by the clinician, then ‘placebo treatment’ is held to be *necessarily* deceptive; that is to say, the deception was necessary for the efficacy of the ‘placebo treatment’[680] (one assumes here that the clinician is indeed deliberately lying and is not either ignorant or negligent). Third, this deception is held to be unethical because it contravenes traditional bioethical principles[681]. As Berger argues, providing ‘placebo treatments’ is supposed to deny the patient the

[676] (H. Brody, 1982) p. 113
[677] (Rorty & Frankel, 2009) p. 17
[678] (Kirsch, 2011) p. 1781
[679] One might object here and argue that the claim that saline is a powerful painkiller is not false in the same straightforward sense as the claim that the saline is morphine – because saline can be a painkiller in the right circumstances. This will be discussed below, but the reply can be anticipated by noting that it is certainly not the saline component which is responsible for any painkilling effects, and therefore the claim is at least somewhat misleading.
[680] This necessity claim is evident in the quotations above.
[681] (Beauchamp & Childress, 2008) Note also that the reasons given for why deception of patients is unethical could of course be recast within any of the broad ‘ethical frameworks’ one might choose. Whether one has deontological or consequentialist intuitions makes no substantive difference.

opportunity to make a fully informed choice about the treatment they receive, and also violate the patient's autonomy (because, according to Berger, the clinician presumes to know that they would accept the clinician's deception, were they informed)[682].

This view has been criticised; usually with the further aim of justifying the use of 'placebo treatments'. There are two broad strategies: there are arguments for the view that such deception is permissible (contrary to what one might expect, based on bioethical principles of respecting autonomy)[683], and there are arguments for the view that deception is in fact not necessarily involved in the use of placebo treatments (circumventing those bioethical principles. That is to say, bioethical principles are irrelevant because there is, at least in some cases, no deception)[684]. The discussion here can be thought of as a species of the latter strategy. While a number of authors have pursued the former strategy, it will be assumed here that deception of the kind illustrated above is indeed unethical – I take it that this is not an implausible assumption to make. My argument will be that even if such deception is unethical, that does not rule out certain kinds of – what some might call – 'placebo treatments'; because they are not necessarily deceptive.

On the view put forward in Chapters 9 and 10, one can deny that 'placebo treatments' are deceptive, simply by noting that there are no such things as 'placebo treatments'. This slightly trivial response highlights that the issue should be reframed in terms of the potential deceptiveness of effective treatments with inefficacious characteristic components instead. It was noted above that there is something disingenuous about such treatments; since one could raise questions about how the inefficacy of the characteristic component might affect the rationale for providing the treatment. Importantly, it is less obvious that such treatments *necessarily* involve deception of patients and are unethical because of this.

Of those authors who have advocated an understanding of 'placebos' and 'placebo effects' that is more sensitive to the research reviewed in Chapter 9, a number of them have gone on to consider the implications this has for the ethical debates about the clinical use of 'placebos'. Howard Brody[685] argues that the clinical use of 'placebo effects' is permissible in so far as it amounts to fostering a

[682] (Berger, 2009)
[683] For example: (Foddy, 2009a, 2009b)
[684] For example: (Moerman, 2002a) (Pittrof & Rubenstein, 2008)
[685] (H. Brody, 2009)

'compassionate, supportive interpersonal relationship[686]' with patients. Maximising the efficacy of these non-characteristic (contextual) components of a treatment is certainly not necessarily deceptive: the clinician does not need to deceive her patients in order to be caring and supportive. Indeed it is argued that maximising these components of treatment represents little more than possessing the skills of a good clinician[687]. In more general terms, it is argued that there is a false dichotomy between full, non-deceptive, disclosure of the details of a treatment on the one hand, and the effectiveness of 'placebo treatments' on the other[688]. That is to say, 'placebo treatments' may be effective by means other than deception.

To elaborate: unethical deception of the sort described at the beginning of this section makes reference only to the apparent efficacy of false beliefs that result from straightforward lies told by a clinician. While one might agree that the cognitive beliefs of a patient can indeed be efficacious, the beliefs that the patient holds about the treatment do not exhaust all of the non-characteristic components of a treatment. Efficacious non-characteristic components need not involve the patient's beliefs ('meaning' is clearly wider than beliefs, but also, for example, in any instance of classical conditioning), nor need those components be in any way related to lies told by the clinician (for example, the colour of a pill and sincere reassurances are clearly not lies on any view[689]).

The point here is that false beliefs possessed by the patient are certainly not the only way that a treatment's non-characteristic components may be effective. Framing the discussion about the ethics of 'placebo treatments' only in terms of false beliefs is simply unsophisticated. This is just to reiterate many of the points made in Chapter 9 and 10, however it is significant because it shows that false beliefs of patients are not necessary for the effectiveness of treatments with an inefficacious characteristic component. Treatments with an inefficacious characteristic component can be effective for reasons that do not involve deception of patients (indeed, even treatments which do have efficacious characteristic components are likely to have

[686] (H. Brody, 2009) p. 13
[687] (Blasi et al., 2001; H. Brody, 2009; F. G. Miller & Luana Colloca, 2009)
[688] (H. Brody, 2009)
[689] Another notable example is Park and Covi's study of 15 'neurotic' patients, who were asked if they would like to receive a sugar-pill containing no medicine – not even a minor deception was perpetrated. They still experienced improvement; as judged by both themselves and the clinicians. Indeed some of the patients refused believe they really did receive a placebo (Park & Covi, 1965)

their effectiveness partly explained by the efficacy of the treatment's non-characteristic components, without any deception of the patients have taken place).

It should be noted however that there are two ways that a treatment with an inefficacious characteristic component may involve the deception of patients. Either the efficacy of the non-characteristic components could depend upon some deception of patients, or the inefficacy of the characteristic component could in some way leads to the deception of patients. The discussion above shows that the former is not the case. Those authors who argue that it is acceptable to maximise the therapeutic effects of the non-characteristic components of a treatment do not therefore provide a full answer to the question of whether treatments with an inefficacious characteristic component are necessarily deceptive. The above shows that it is permissible to exploit the efficacy of at least some of a treatment's non-characteristic components, namely those that do not entail any deception in order to be efficacious (since there are many components of a treatment which may be efficacious, and which also are not related to the patients cognitive beliefs, then there surely are such components whose effects can be non-deceptively maximised – take the empathy of the clinician as a paradigm case of such a non-characteristic component). Showing that one can, without the deception of patients, exploit the efficacy of a treatment's non-characteristic components does not however entail that one can, without deception, provide treatments with ineffective characteristic components. Further arguments are needed to decide whether providing effective treatments with inefficacious characteristic components leads to the deception of patients.

This latter issue is more problematic. The problem facing treatments that are known to have inefficacious characteristic components would seem to be that patients can be misled in more subtle ways besides being deceived by straightforward lying by the clinician. Against the view that giving 'placebo treatments' must necessarily involve deceiving patients, consider that Lichtenberg et al claim that the following, entirely true, statement could accompany the (apparently ethical) use of a treatment with an inefficacious characteristic component:

> 'I would like to offer you a pill which I believe can help lessen your suffering. I do not know exactly how it works. I have other pills to offer whose mechanism is clearer, but I am not sure that

> they will work better for you, and they may also entail more serious side effects[690]'

From which Lichtenberg et al conclude:

> 'In this manner, the physician is being open and honest with the patient[691]'

Statements such as this attempt to carve out a non-deceptive role for 'placebo treatments', however they possess an air of ethical double-speak. It is unclear whether providing, say, sugar pills, along with the information above involves clinicians deceiving their patients or not. In an attempt to capture the subtle way in which this approach to the provision of 'placebo treatments' seems deceptive a number of authors have argued that one can mislead merely by manipulating the albeit true information given to patients[692]. Their general point is that only making true statements is not sufficient to avoid misleading patients. Brody for example quotes Richard Cabot, writing in 1903, on the same point: 'a true impression, not certain words literally true, is what we must try to convey[693]'. To put the point more concretely, Brody also illustrates the tacit expectations patients are warranted to form about their treatments, as follows:

> 'if a drug or other treatment is given, it is selected for its pharmacologic potency for the patient's condition. It also seems reasonable to assume that the patient will not expect that the physician will specifically name the treatment—the patient is accustomed to receiving pills alluded to by the physician merely as "an antibiotic" or "a decongestant," but these remedies are still assumed by the patient to be pharmacologically potent[694]'

---

[690] (Lichtenberg et al., 2004) p. 552 Similar 'acceptable advice' statements occur in, for example: (Park & Covi, 1965) (Pittrof & Rubenstein, 2008) (F. G. Miller & Luana Colloca, 2009)
[691] (Lichtenberg et al., 2004) p. 552
[692] (Hester & Talisse, 2009; Schwab, 2009) In a similar vein, see also, concerning the Additivity assumption in a research context: (F. G. Miller & Luana Colloca, 2011)
[693] Cabot quoted in (H. Brody, 1982) p. 114.
[694] (H. Brody, 1982) p. 115

The idea here is that there is an expectation that if one receives a treatment, then one is receiving it because its characteristic component is efficacious: a pill referred to as an antibiotic is expected to be pharmacologically potent. Deception arises from the fact that a clinician can claim, truly, that a treatment – such as 'an antibiotic' – may be effective for a patient's viral infection whilst implicitly denying that any claims about the efficacy of specific components are entailed by the effectiveness of the treatment overall. To put this another way, deception arises from the fact that a clinician can make true claims about the effectiveness of a treatment, whilst it also being the case that they know the characteristic component of the treatment is inefficacious, and crucially whilst allowing the patient to form warranted beliefs about the efficacy of the characteristic component.

When it is known that the characteristic component is inefficacious I claim the clinician is acting disingenuously and deceptively if they give the patient information of the kind quoted above. This is an instance of the general fact that what one knows puts constraints on what is permissible[695]. As Brody argued, being given some particular treatment tacitly implies that the characteristic component of that treatment is efficacious for the condition being treated. That implication stems from the fact that the characteristic component is a value-laden concept; it is intertwined with the rationale for giving the treatment. A treatment with an inefficacious characteristic component subverts that rationale and, I claim, deserves some special explanation by the clinician.

The problem with effective treatments which have inefficacious characteristic components is that there is greater scope for clinicians to convey a false impression about the source of their effectiveness. The reason is that, as claimed above, it is reasonable to expect the characteristic component to be efficacious. In fact, I suggest that this idea can be generalised. Just as a clinician who provides a patient with 'an antibiotic' creates the expectation that the antibiotic (that is, characteristic) component of the treatment is efficacious; so the mere provision of a pill warrants the assumption that the characteristic component is pharmacological. Indeed, as noted in Chapter 10, the reason one gives exactly similar non-drug-containing pills to placebo groups in PCTs of a drug is precisely because one wants to create, in the placebo group, the expectation in patients that they are receiving a pharmacological treatment.

[695] (Worrall, 2008)

On the basis of the above I suggest that the efficacy of the characteristic component matters because it underlies the rationale and permissibility of providing the treatment. When the characteristic component of a treatment is known to be inefficacious, providing that treatment seems more likely to mislead and deceive the patient. In such cases the explanation of the treatments effectiveness subverts one's reasonable expectations. The rationale for providing the treatment is unintuitive, since one would typically expect a treatment to be provided because the component which defined it as that treatment was efficacious for the condition being treated. This unintuitive rationale makes it more likely that patients will be deceived. This is ethically problematic for the traditional reasons that deception is supposed to be problematic: the patient makes treatment choices based on false information, and the clinician violates their autonomy by denying them information on which to make treatment choices[696].

Notice most importantly that the argument developed above provides a reason to believe that treatments with inefficacious characteristic components are, other things equal, more prone to involve some deception of patients than if the characteristic component were efficacious. It does not provide a reason to believe that they necessarily involve the deception of patients. Whether any deception takes place is a matter of the treatment context; not, as is assumed in arguments about the ethics of 'placebo treatments', simply a fact that follows from the treatment being a 'placebo'. Consequently, I now suggest that deception of patients is avoidable when, for example, those patients are properly informed about the inefficacy of the characteristic component of their treatment.

With the similar aim of ameliorating the deceptive nature of 'placebo treatments' Pittrof and Rubenstein[697], as well as Lichtenberg[698] have offered a set of necessary and jointly sufficient criteria for the ethical provision of 'placebo treatments'. Their respective criteria, I claim, are insufficient to justify the use of treatments with an inefficacious characteristic component, however. In both cases Pittrof and Rubenstein, and Lichtenberg's criteria can be summarised into two general kinds of condition. Firstly, both involve an effectiveness condition: The

[696] As noted above, the link between deceptive treatments and unethical treatments is assumed in this discussion, and not examined further.
[697] (Pittrof & Rubenstein, 2008)
[698] (Lichtenberg et al., 2004)

treatment ought to be effective and not merely mollifying[699], and also, there ought not to be any alternative 'gold standard' treatments that are more effective[700]. Second both involve an information condition: The treatment ought only to be given to patients who are fully informed about the nature of the treatment they are receiving and consent to receiving it. Both note however that full disclosure of the evidence-based for the treatment should occur 'if they [the patients] ask[701]', or 'when asked[702]'.

I claim that the information condition is insufficient to ameliorate the deception that may occur when a treatment with an inefficacious characteristic component is given. Indeed, the statement that Lichtenberg suggests should accompany the 'placebo treatment' is precisely that statement quoted above, as an example of advice that is likely to be misleading. In contrast then, I suggest that the more subtle kinds of deception, which treatments with inefficacious characteristic components are likely to lead to, must be explicitly addressed. Importantly, the fact that the characteristic component is inefficacious ought to be made clear, without the patient needing to ask a question for that information to be revealed.

As will be shown in Part Four, this argument has significant implications for the ethical arguments that are made against the provision of homeopathic treatments.

11.3 Summary

The purpose of this chapter was to draw out further and more fully the implications of the evidence reviewed in Chapter 9 and the argument made in Chapter 10.

Firstly, how should one re-evaluate the distinction between efficacy and effectiveness if one is to abandon reference to 'placebos' and 'placebo effects'?

The results from the preceding chapters allow one to dismiss, for example, Walach's 'paradox of efficacy'; doing so highlights the importance of being precise about which particular component one has in mind when one claims, for example

[699] (Lichtenberg et al., 2004) p. 553
[700] (Pittrof & Rubenstein, 2008) p. 1020
[701] (Pittrof & Rubenstein, 2008) p. 1020
[702] (Lichtenberg et al., 2004) p. 553

that “this treatment is efficacious”. Most significantly the term ‘characteristic component’ was introduced as a way of talking about the component that defined a treatment as that treatment. It was argued that the distinction between efficacy and effectiveness is not substantive; it is merely a useful way to denote two different comparisons that one might perform. It was also argued that this was independent of other ways that the term effectiveness might be used, to convey clinical usefulness or to convey facts about the robustness of treatment’s (or some component’s) efficacy across different circumstances.

Secondly, the notion of a characteristic component was examined further. In particular, why does the efficacy of the characteristic component matter?

It was argued that the efficacy of the characteristic component matters for ethical reasons. It seems there is something disingenuous about providing patients with treatments that have an inefficacious characteristic component. Whilst this idea is clearly related to the idea that ‘placebo treatments’ are necessarily deceptive, restating that idea without reference to ‘placebos’ makes the necessity of the deception harder to sustain. It was suggested that treatments with inefficacious characteristic components could be provided ethically, however the information given to patients must be carefully managed; it was argued that simply making true claims about treatments with inefficacious characteristic components was insufficient; rather the reason why such treatments are effective ought to be explicitly explained to patients, in order to avoid deception.

CHAPTER 12

12. Summary of Part Three

Part Three presents a rather different picture of 'placebos' and 'placebo effects' than is seen in the homeopathy controversy, or one might intuitively hold.

In Chapter 9 the research literature about placebos and placebo effects was reviewed. I argued that this research speaks against the view that placebo effects are merely psychological phenomena or that they point to problems in the biomedical paradigm. On the contrary, placebo effects are the result of a wide range of factors, which act through specific physiological mechanisms: there is no single 'placebo effect'; and there are multiple mechanisms by which such effects are generated. The most coherent attempt to understand placebo effects conceptualises them in terms of the meaning and significance that treatments and the treatment context has for patients.

In Chapter 10 the role of 'placebos' and 'placebo effects' in a research context was questioned, and developing the ideas in Chapter 9 a little further, it was argued that the terms should be abandoned. I argue for abandoning the terms 'placebo' and 'placebo effects' because they serve no analytical purpose. It is a mistake, I argue, to think of placebo comparison as a case where something is compared to 'a placebo'. Rather, placebo comparison should be understood as a situation which sets-up the treatment and control groups in a particular way; not as a case involving objects or procedures called 'placebos' employed in order to control for 'placebo effects'. The meaningful questions to ask involve being specific about the details of the controls – so that one can evaluate the plausibility of alternative explanations of the results (See Part Two). One has a better view of what is going on in a placebo comparison if our descriptions don't use the terms 'placebo' or 'placebo effect', they obscure legitimate questions about the specific details of the control group.

In Chapter 11 the implications of the evidence reviewed in Chapter 9 and the argument made in Chapter 10 were drawn out more fully. Firstly concerning the distinction between efficacy and effectiveness, secondly concerning the question of why the efficacy of the characteristic component seems to matter. The first discussion clarified two different ways that the distinction could be drawn, and more

importantly, introduced the term 'characteristic component'. The second discussion argued that the efficacy of the characteristic component matters for ethical reasons. The issues are similar to previous work looking at the ethics of placebo treatments. However contrary to the common idea that 'placebo treatments' are unethical, it was suggested that treatments with inefficacious characteristic components could be provided ethically, although the information given to patients must be carefully managed because it is easier for patients to be misled about the effectiveness of their treatments.

12.1 Introduction to Part Four

Part Four attempts to integrate the findings from Parts Two and Three by applying them to the homeopathy controversy discussed in Part One. The question posed at the end of Part One, which was the organising question for this thesis, was to what extend the concepts of EBM and 'placebo' provides a solid foundation for the Canonical Criticism of homeopathy. Part Four addresses this directly. It focuses on how the five key ideas in the Canonical Criticism (See Part One, and also Chapter 13) can be reinterpreted in light of the arguments put forward in Parts Two and Three.

With regard to the evidential debate in particular, Part Four firstly evaluates whether the Canonical Criticism's interpretation of EBM is acceptable, and proposes to resolve the tension noted previously between the STC's dismissal of mechanistic evidence and the Implausibility Argument made in the wider literature against the view that homeopathic treatments could be efficacious. It evaluates secondly the special role that placebo comparison plays in the debate about homeopathy; seeking also to integrate the arguments about 'complexity' made by proponents of homeopathy in an answer to the question of whether placebo comparisons of homeopathic treatments are possible.

With regard to the policy debate, Part Four firstly evaluates whether the Canonical Criticism uses 'placebos' as an acceptable ethical standard. It provides a reinterpretation of the No Placebos argument described in Part One. Part Four does not address the Indirect Harm argument, which was the second of the ethical arguments marshalled by opponents of homeopathy in the policy debate. However, Part Four does discuss whether there is a possible role for homeopathic treatment in

healthcare, in circumstances where one should be less worried about any potential harms.

This whole discussion takes place in Chapter 13. Chapter 14 provides a summary of the conclusions of the thesis.

PART FOUR: RE-EVALUATING THE CONTROVERSY

CHAPTER 13

13. The Canonical Criticism revisited

In Part One five key points in the Canonical Criticism were noted, they were:

Evidential debate

(1) Evidence-based medicine provides the framework for assessing whether homeopathy works. It is a question of efficacy: do homeopathic treatments outperform placebo in randomised trials.
(2) The best available evidence (from randomised trials, or better, meta-analyses of such trials) shows that homeopathic treatments are equivalent to placebo.
(3) The homeopathy=placebo hypothesis is supported by mechanistic evidence which shows that it is implausible to expect homeopathic treatment to be efficacious.

Policy debate

(4) No Placebos argument: The provision of placebo treatments (and therefore homeopathy) necessarily involves deceiving, or violating the autonomy of, patients; and also contributes to the medicalisation of the patients' complaints.
(5) Indirect Harm argument: The provision and state endorsement of placebo treatments (and therefore homeopathy) causes Indirect Harm in so far as it creates the perception that they are efficacious medicines, because this perception may delay the treatment of serious conditions, or undermine public health advice.

The structure of the argument which the Canonical Criticism built around this began with the claim that homeopathic medicines are not efficacious; rather they are placebos. Opponents of homeopathy note that the inefficacy of homeopathic treatment does not rule out that they can appear to be efficacious, through a combination of placebo effects, and the natural progression of a condition (for example, the condition may be self-limiting or vary in severity). It is the fact that homeopathic treatment can appear to be efficacious which necessitates an assessment of homeopathy in PCTs. The key contentions are that homeopathic

treatment does not come out favourably in PCTs, and that, for this reason, it is impermissible to give placebo treatments to patients, even if they are effective.

Parts Two and Three supply the conceptual tools with which to examine points (1)-(5) and the argument above. From Part Two, the EBM philosophy of evidence was specified in such a way that it is not tied to 'evidence hierarchies' and favours instead the view that any evidence can be good evidence if it discriminates between hypotheses. From Part Three the use of the terms 'placebo' and 'placebo effect' can be abandoned. I suggest that both of these ideas allow one to talk with greater precision about what it means to claim that a treatment works, and illuminates the ethical debate about 'placebo treatments'.

In what follows I put forward some of the key conclusions that these ideas allow one to draw about the homeopathy controversy. Specifically in response to the questions posed in Chapter 4, and to prefigure this chapter I argue that: (§13.1.1) in relation to the evidential debate I claim that the STC report undervalues mechanistic evidence, on account of their commitment to a Categorical Interpretation of EBM. I also claim that (§13.1.2) if, as proponents of homeopathy assert, Additivity fails in the case of homeopathic treatments, this does not support the view that the quasi-pharmacological component of the treatment is the characteristic component nor the view that its efficacy cannot be meaningfully measured in placebo controlled trials. In relation to the policy debate I focus solely on the ethical arguments put forward in the Canonical Criticism. I claim that (§13.2.1) the No Placebos argument fails and tentatively suggest (§13.2.2) circumstances in which the provision of homeopathic treatment would seem to be permissible, even given the Canonical Criticism's view about the evidential debate.

13.1 The evidential debate about homeopathy

There are at least two issues with (1)-(3) above that I wish to highlight. The first concerns a tension between (1) & (2), and (3), relating to the weight given to mechanistic evidence. The second concerns a presupposition of (1), relating to the characteristic component of homeopathic treatment.

13.1.1 The House of Common's Science & Technology Committee and the Implausibility Argument

Proponents of homeopathy often single out the concept of 'evidence-based medicine' for criticism, because it is a concept on which the Canonical Criticism draws heavily. Potential problems arise when one considers the particular interpretation of EBM that is offered in the Canonical Criticism.

In Part One it was shown that proponents of homeopathy argue that although the medical literature talks about EBM involving the integration of different kinds of evidence, the EBM philosophy of evidence – especially in debates about controversial treatments, like homeopathy – is one dimensional and unsophisticated. The Canonical Criticism is accused of reifying evidence from randomised trials. The same criticism is also made from within the medical literature; advocates of EBM are accused of holding a Categorical Interpretation of EBM, meaning that evidence from randomised trials is thought to 'trump' all other kinds of evidence.

This reveals a tension in the way that opponents of homeopathy construct the evidential debate. The STC exemplify this tension most, because unlike other sources of the Canonical Criticism, they explicitly deny that mechanistic evidence should play a role in their evaluation of homeopathy. If the critics of EBM are correct, then there is a problem in the way that (1) & (2) are combined with (3). It is unclear how (3) can possess significant evidential weight, if (1) & (2) are taken seriously. That is to say, opponents of homeopathy seem committed to an interpretation of EBM that holds that mechanistic evidence possesses little evidential weight, but also assert that homeopathic treatments cannot work because they have a grossly implausible mechanism (The STC are at least consistent in their application of a Categorical Interpretation of EBM; in so far as they deny that mechanistic considerations possess evidential weight). Whether this is a genuine tension, and the way to deal with that tension, is suggested by the results from Part Two.

Part Two asked firstly whether this Categorical Interpretation of EBM really does represent how EBM is interpreted in the medical literature. The purpose of asking this question was to evaluate the extent to which EBM provides an adequate foundation for the arguments put forward in the Canonical Criticism. If EBM is not interpreted categorically, then the criticisms based on the notion that it is are misplaced (Although on any interpretation of EBM one would like some account of how the different kinds of evidence marshalled in the Canonical Criticism should be

consistently combined). If on the other hand EBM is interpreted categorically, then the challenge made by proponents of homeopathy and other critics of EBM undermines the Implausibility Argument made in the Canonical Criticism.

It was shown in Part Two that the EBM literature is unhelpfully unclear about how EBM is interpreted. The examination of the medical literature presented in Chapter 5 showed that there was a set of basic arguments for EBM from which only very weak conclusions were drawn. This was developed more systematically in Chapter 6. A large corpus of papers about EBM were analysed and it was shown that there was no clear 'EBM view' reflected in the literature. I claimed further that the EBM literature looks confusing precisely because the concept is confused: the EBM literature is simply unclear, it is not the case that essence is hidden in noise. Whereas other authors have suggested that there is widespread misunderstanding of EBM, I claim there is flexibility of interpretation. This reinforced the need for critical clarification of what the EBM view should be. Nearly twenty years of literature has been surprisingly unhelpful in answering this question.

Part Two also therefore briefly examined the question of what interpretation *should* be held. The Categorical Interpretation stems from giving EBM's evidence hierarchies an epistemic reading. That is to say, reading evidence hierarchies as providing an epistemological template that determines the level of evidential support that different research designs give to medical claims. By drawing on recent work by philosophers of science, it is argued that the Categorical Interpretation is not a defensible interpretation of EBM. On a better interpretation, suggested by John Worrall and Jeremy Howick, it is argued that EBM is epistemologically unexceptional. Furthermore it was argued that evidence hierarchies should be interpreted heuristically, meaning that although they could be used as an aid to busy clinicians, they do not possess any deep epistemological significance. Outside of resource restricted circumstances there is no substitute for the hard work of critically appraising the total evidence. One important consequence of the view put forward by Worrall and Howick is that any kind of evidence may potentially offer support to a hypothesis. This idea provides the tools for resolving the tension between (1) & (2), and (3). This is the first conclusion:

Conclusion 1

There is no single or stable interpretation of EBM in the medical literature. The literature is, and always has been, unclear about what the details of EBM amount to. However the most straightforward interpretation, the Categorical Interpretation, is not defensible.

The STC Evidence Check report clearly endorses a Categorical Interpretation of EBM. It explicitly states that only evidence from randomised trials (or better, from meta-analyses of such trials) is appropriate for evaluating homeopathic treatments[703]. Elsewhere in the STC report[704] and also in the wider literature[705] however, the Canonical Criticism includes the, often repeated, Implausibility Argument. This implies that a more prominent evidential role is being given to mechanistic reasoning than is warranted on a strict Categorical Interpretation.

Note that I do not propose to evaluate the mechanistic evidence. I intend only to comment on the way that mechanistic evidence is used in the homeopathy controversy. Note second that a preliminary refinement of the Implausibility Argument is necessary before the discussion proceeds. In Part One the Implausibility Argument was taken to be a claim about the implausibility (if not impossibility) of a mechanism by which homeopathic treatments could have therapeutic effects. The Implausibility Argument is more precisely a claim about the mechanism by which one particular component of homeopathic treatment could have therapeutic effects on a given condition. That is, the effect of the contents of homeopathically prepared pills – call this component the 'quasi-pharmacological component' of homeopathic treatment. The phrase 'quasi-pharmacological' is not meant to prejudice the discussion, but rather to indicate that the content of homeopathically prepared pills is unconventional. It is unconventional simply because it has been prepared in accordance with the small doses and dynamisation principles of homeopathy (see Chapter 2) – it is not controversial to state that the contents of homeopathic pills are not what would conventionally be thought of as pharmacological content. It is the efficacy of this quasi-pharmacological component of homeopathic treatment that the Implausibility Argument aims to refute.

[703] (House of Commons Science & Technology Committee, 2010) paras 19-26
[704] (House of Commons Science & Technology Committee, 2010) paras 48-64
[705] (Baum, 2006; Baum & Edzard Ernst, 2009; Edzard Ernst, 2007, 2011a, 2011b; Holt et al., 2011; Pandolfi, 2010, 2011; Sehon & D. Stanley, 2010; D. Stanley & Sehon, 2011)

Caveats and terminology aside, there is a key difference in the way that the STC report deals with the Implausibility Argument, and the way that it is dealt with in the wider critical literature. The STC set aside their judgement about the implausibility of homeopathy. Although the STC claim that there is indeed no plausible way the homeopathic principles of similarity and small doses could account for the efficacy of the quasi-pharmacologic component of homeopathic treatment, they do not make use of this point in their report, instead they explicitly state: 'while we comment on explanations for how homeopathy works, it is not a key part of our Evidence Check[706]'. I suggest this is a mistake: one which arises from their holding a (indefensible) Categorical Interpretation of EBM.

Contrary to the STC's view, the argument developed in Part Two showed that any evidence could be counted as good evidence if it was able to discriminate between plausible alternatives. Mechanistic evidence is no exception (even though, as Howick has argued[707], there is seldom the requisite knowledge to make reliable inferences from mechanisms to therapeutic claims). In the case of the Implausibility Argument the form of the inference is different from that which is typically considered, however. Howick for example is concerned with the inference from knowledge of a mechanism to knowledge of a causal link[708]. The Implausibility Argument makes the inverse of this inference: from knowledge that there is no mechanism to the knowledge of the impossibility of a causal link. More specifically the argument is that the quasi-pharmacological component of homeopathic treatment doesn't cause therapeutic effects, because there can't be a mechanism by which it could cause therapeutic effects[709]. Using the insights from Part Two, what can be said about this form of inference?

The claim that the Implausibility Argument makes is one which, following Part Two, is clearly highly discriminating between rival hypotheses. The mechanistic evidence against the efficacy of the quasi-pharmacological component of homeopathic treatment is good evidence if and only if the Implausibility Argument can justify the claim that there can be no mechanism by which it could be efficacious.

---

[706] (House of Commons Science & Technology Committee, 2010) para 18

[707] (Howick, 2011) Ch. 10

[708] (Howick, 2011) Many other philosophers have also addressed the relationship between mechanisms and causation, these issues are not immediately relevant to this discussion, however see: (Glennan, 1996, 2002; Machamer et al., 2000; Russo & J. Williamson, 2007)

[709] Clearly this relies on the ontological claim that causal links can be explained mechanistically: to reiterate, these issues are not addressed here. See: (Glennan, 1996, 2002)

If the Implausibility Argument achieves what it purports to, then one has excellent mechanistic evidence against the efficacy of the quasi-pharmacological component of homeopathic treatment.

The Implausibility Argument rests on a very strong empirical claim (there can be no mechanism). In fact this might seem dogmatic to the extent that one might question whether it is truly the claim that opponents of homeopathy are making. I would argue that this really is the claim being made. The key question to ask is whether the critics of homeopathy would admit the possibility that there could be a mechanism, Baum and Ernst are clear: 'we think that a belief in homeopathy exceeds the tolerance of an open mind. We should start from the premise that homeopathy cannot work[710]'. Now it may be that Baum and Ernst are correct. Undeniably, there are views that it is pointless to engage with dialectically[711]. Of course the key question is whether the Implausibility Argument really does justify the claim that there can be no mechanism by which the quasi-pharmacological component of homeopathic treatment can work.

It seems puzzling that the STC believe that the Implausibility Argument does justify such a claim, but do not marshal this evidence. The STC hold an interpretation of EBM from which they infer that 'lack of scientific plausibility is disappointing, but does not necessarily mean that a treatment does not work[712]'. Indeed, the STC endorse the key empirical premise of the Implausibility Argument; namely, that the principles upon which homeopathic treatments are prepared rules out the efficacy of the treatment's quasi-pharmacological component. However, despite the fact that they claim a mechanism is indeed implausible, they hold the view that this has no consequences for assessing efficacy.

On the contrary however, I claim that given the strength of the claim the STC endorse about the mechanistic evidence, then it ought to be a part of their 'Evidence Check'. The Categorical Interpretation of EBM, and the interpretation put forward by Worrall and Howick in Part Two diverge when there is strong mechanistic evidence, of precisely the sort that it is claimed there is by the STC and by opponents of homeopathy more widely.

---

[710] (Baum & Edzard Ernst, 2009)

[711] As Timothy Williamson notes: 'by accepting the dialectical standard [of evidence] unconditionally, we lay ourselves open to exploitation by ruthless opponents… When one is warranted in refusing to play the sceptic's dialectical game, the dialectical standard of evidence becomes irrelevant' (T. Williamson, 2007) p. 238-9

[712] (House of Commons Science & Technology Committee, 2010) para 65

If one were to be generous to the Canonical Criticism, one might argue that the fact that mechanistic evidence can be good evidence is indeed acknowledged, as demonstrated by the fact that the Implausibility Argument is made at all. However, why then does the Canonical Criticism also hold a Categorical Interpretation of EBM? The STC's view is perhaps the most consistent statement of the Canonical Criticism, however it is deficient. The STC's position is deficient in so far as it ignores mechanistic evidence. The problem, as diagnosed above, is that the STC hold a view about the EBM philosophy of medicine that incorrectly assigns a weak evidential role to mechanistic reasoning. To reiterate: given what the STC claim is true about the plausibility of the principles of homeopathy, it follows from Part Two that they should have made a stronger argument on that basis.

The second and third conclusions that I wish to draw are therefore as follows:

Conclusion 2

There are no a priori constraints on what kinds of methods can generate good evidence (From Part Two). Any evaluation of the evidence for the efficacy of the quasi-pharmacological component of homeopathic treatment ought to take into account the mechanistic evidence for and against its purported efficacy.

Conclusion 3

The House of Common's Science & Technology Committee undervalue mechanistic evidence because they hold a Categorical Interpretation of evidence-based medicine. This is of particular significance because, in fact, they endorse strong claims about the mechanistic evidence against the efficacy of the quasi-pharmacological component of homeopathic treatment.

Proponents of homeopathy marshal a number of different kinds of mechanistic evidence in their counter-arguments to the Implausibility Argument; their claim was simply that a mechanism for the efficacy of the quasi-pharmacological component of homeopathic treatments was possible[713]. This will not

[713] This claim rested on evidence from materials science about the physical-chemistry of water, as well as evidence from laboratory research on the action of high dilutions

be discussed in detail; however it is worth briefly noting a point about the aim of proponents' counter-arguments:

Consider that proponents of homeopathy are arguing in the more typical way, from mechanistic knowledge to justified belief in a causal link. As was noted in Part Two, and referred to above, mechanistic evidence *can be*, but seldom is, good evidence[714]. Proponents of homeopathy do not make claims about the mechanistic evidence that are as strong as those made by opponents. Indeed, the incompleteness of the mechanistic evidence for the efficacy of the quasi-pharmacological component is not a controversial point. To reach that conclusion one needs only to claim that those mechanisms are understood to a lesser (or at best, similar) degree as mechanisms for conventional medicines. In general, one shouldn't rely on biological theory as evidence for the efficacy of the pharmacological components of conventional treatments: that mechanistic evidence often fails to be a reliable guide to therapeutic benefit is something the Categorical Interpretation and Worrall and Howick's view, put forward in Part Two, agree on (they disagree that it is *always* weak). Given this then, a fortiori, one shouldn't rely on knowledge of mechanisms as evidence for the efficacy of the quasi-pharmacological components homeopathic treatments; since the mechanistic knowledge supporting conventional treatments is substantially greater and yet still often insufficient.

As a result one might argue that proponents of homeopathy aim only to prevent the efficacy of the quasi-pharmacological component of homeopathy being ruled out, tout court; rather than aiming to provide a complete mechanistic model for the efficacy of the quasi-pharmacological component. It seems unlikely that the counter-arguments made by proponents of homeopathy provide good mechanistic evidence for the efficacy of the quasi-pharmacological component of homeopathic treatment. Equally however it seems unlikely that this is the main purpose of those counter-arguments; rather, I would argue that the aim is simply to put the mechanistic debate 'on the table'. That is, to show that unless the Implausibility Argument can justify its strong claims, it too is not good mechanistic evidence.

Two conclusions have been drawn about the way that the evidence is used in relation to determining the efficacy of the quasi-pharmacological component of

---

on animal cells in vitro (and perhaps also included evidence provided by analogies with the mechanisms of vaccines and hormesis) – See Part One for references.

[714] (Howick, 2011) Ch. 10

homeopathic treatment. A hidden premise of this discussion is that the quasi-pharmacological component matters because it is truly the characteristic component of homeopathic treatment. This is the second issue that will be addressed.

13.1.2 The characteristic component of homeopathic treatment

In Part One it was shown that the evidential debate is framed in terms of whether or not homeopathic treatments are equivalent to 'placebo treatment'. The Canonical Criticism generally, and the STC Report in particular, attach importance to this equivalence because it draws the boundary between efficacy and effectiveness. The argument put forward by opponents of homeopathy was that efficacy was only demonstrated when a treatment could outperform placebo. Mere effectiveness, equivalent to placebo, was judged by opponents of homeopathy to be insufficient to claim that homeopathic treatment worked, because it could be effective for reasons that had nothing to do with the action of the homeopathic medicine specifically. Framing the evidential debate in terms of (1) demonstrates that the debate is about why homeopathic treatment is effective, rather than simply whether it is effective.

In Part Three, the use of the concept of 'placebo' as an evidential and ethical standard was examined. The view developed in Part Three highlights the vagueness of claims that a treatment is equivalent to placebo. If claims are being made about treatments, then many different placebo comparisons involving that treatment are possible – the questions to ask of course are which component is being singled out, and whether the placebo group is appropriately set-up to ensure only that component is singled out. Consequently, for any use of the term 'equivalent to placebo', more needs to be said about what comparison one has in mind. Significantly, the notion of a treatment's characteristic component was introduced, following Grünbaum, in order to capture the component that one would most likely have implicitly in mind when talking about a placebo comparison involving a given treatment. Drug treatments illustrate the idea of a characteristic component well: for example, the drug paracetamol is the characteristic component of paracetamol treatment for pain relief. It was then shown that the efficacy of the characteristic component is important for ethical reasons; it is intertwined with the rationale for providing that treatment rather than some other.

On this view it follows that 'homeopathy is equivalent to placebo' should be read as the claim that the characteristic component of homeopathic treatment is inefficacious. Reinterpreting the claim without reference to placebo adds precision; it also alters the way that the premise 'homeopathy is equivalent to placebo' can be used in ethical and policy arguments.

The Canonical Criticism holds the view that the characteristic component of homeopathic treatment is the quasi-pharmacological component. It is argued that the most important claim for drawing ethical and policy conclusions about homeopathic treatment is the inefficacy of its quasi-pharmacological component. On this view the most important fact about homeopathic treatment is whether, when receiving treatment from a homeopath, it would make a difference if one were given a sugar pill or the homeopathic pill. This fact is only indirectly related to the overall effectiveness of homeopathic treatment, but directly related to the explanation of why it is effective. The importance of this fact rests on an ethical argument: opponents of homeopathy claim that it is crucial that the characteristic component of the treatment should form part of the explanation of why homeopathic treatment is effective, otherwise the treatment will deceive patients (the No Placebos argument, from Part One) and, or, result in various kinds of harm (the Indirect Harm argument, from Part One). A simple framing of the homeopathy controversy as asking whether homeopathic treatments work, and answering with reference to the clinical research literature misses the more fundamental point that such a question is irrelevant without some account of why the efficacy of the characteristic component matters. Such an account should not, from Part Three, refer to 'placebos' to do any of the evidential or normative work. Note that this is a different and more complicated position that opponents of homeopathy must defend (which will partly be returned to in §13.2). This gives the fourth conclusion:

Conclusion 4

When evaluating whether homeopathic treatment 'works', the key concern is with the efficacy of the characteristic component, but the efficacy of the characteristic component is only important for ethical reasons. Opponents of homeopathy who claim it does not 'work' must be seen as expressing an ethical objection to the reasons why it is effective.

The challenge made by proponents of homeopathy to the validity of placebo controls can be read as a challenge to the account, given in the Canonical Criticism, that the quasi-pharmacological component of homeopathic treatments is, in fact, the characteristic component. Proponents of homeopathy challenge the, in principle, testability of homeopathic medicines in randomised trials. Such challenges are premised on some notion of the complexity of homeopathic treatment. They aim to show that it is illegitimate to attempt to single out specific components of the treatment, and thereby question the validity of placebo-comparison as an evidential standard, applicable to homeopathic treatments. This is a challenge to the notion that a placebo comparison that singles-out the quasi-pharmacological component of homeopathic treatment can illuminate why homeopathic treatment works.

I suggest that, although ultimately unsuccessful, there is more to be said about this challenge. The arguments from Part Three allow the debate about the in principle testability of homeopathic treatments in PCTs to be described in more rigorous terms. They also suggest ways that the validity of those arguments could be investigated empirically; and suggests some constraints on what counts as a legitimate PCT of homeopathic treatments.

Consider two arguments one might make, on the basis of Part Three, in order to illuminate points made by proponents of homeopathy who question the validity of placebo controlled trials of homeopathy. Both arguments object to singling-out the quasi-pharmacological component and they both make the claim that this is an illegitimate test of the efficacy of the characteristic component; as follows: First, it could be argued that the quasi-pharmacological component is only one part of the characteristic component of homeopathy. Second, it could be argued that there are problems with assuming Additivity in the case of homeopathic treatment. In both cases I argue that these considerations do not present any fundamental problem to the validity of placebo-comparison as an evidential standard, applicable to homeopathic treatments.

In the first case, an analogy can be drawn between homeopathic treatment and polypill treatment[715]. This would involve making the claim that the characteristic component of homeopathy should include other aspects of the treatment; much in the same way that the characteristic components of a ‘polypill’ is, quite legitimately,

[715] See Chapter 11 for discussion of polypills

made up of different drugs[716]. That is, just as the efficacy of one of the drug components of a polypill is not equivalent to the characteristic component of the polypill but only one part of it, so the efficacy of the quasi-pharmacological component of homeopathic treatment is only one part of the characteristic component of homeopathic treatment. Just such an argument is made by Thompson et al, who suggest homeopathy is a 'complex intervention'. They state in their conclusion:

> 'the consultational activity within homeopathic care has aspects which are specific to homeopathy. If these aspects are therapeutically active, which is a reasonable working hypothesis, then comparison of placebo and non-placebo arms in homeopathic trials will not constitute a fair test. This is because the patients in the placebo arms will be receiving an active and specific part of the homeopathic care[717],

Opponents of homeopathy are likely to make the following reply to this argument: in the polypill case one knows that each component is individually efficacious, whereas the quasi-pharmacological component is not efficacious, according to the Canonical Criticism. If therefore there is a set of components of homeopathic treatment which are jointly efficacious, and which can justifiably be called the characteristic component, then it would seem that the quasi-pharmacological component of the treatment is at best a redundant member of this set. A better analogy therefore would be between the quasi-pharmacological component of homeopathic treatment and one of the inefficacious excipients used in the polypill.

More fundamentally, and contrary to the claim of Thompson et al, treating homeopathic treatment as 'complex' in the sense that the characteristic component can be broken down into further components (like a polypill) does not seem to present a challenge to the idea that one could perform placebo comparisons on each

[716] In the example cited in Chapter 11, the PILL Collaboration study, the polypill used was made up of four component drugs (aspirin, lisinopril, hydrochlorothiazide and simvastatin), and it was the combined efficacy of these four drugs that was taken to be the characteristic component of the polypill treatment.
[717] (T. D. B. Thompson & Weiss, 2006) p. 14

component of the treatment's characteristic component. The fact that the characteristic component of the polypill is made up of different drug components does not speak against investigations of the efficacy of those drug components individually. If the characteristic component of homeopathic treatment is made up of other components besides the quasi-pharmacological component, then that does not speak against investigations of the efficacy of that component individually.

Thompson et al's argument seems to be that placebo controlled trials of the quasi-pharmacological component are not 'fair tests' of homeopathic treatment, because other components are efficacious too. Of course, that depends on what claim one is seeking to investigate. This is not to say that there are no efficacious components of homeopathic treatment; surely, there are. Rather the key point turns on what it is that proponents of homeopathy claim about the treatment. The principles which underlie homeopathic treatment are not necessary for explaining why many of the other components of homeopathic treatment are likely to be efficacious. As Part Three suggests, the dynamics of patients' expectations and the long and involved consultation process, for example, are likely to be part of the explanation why homeopathic treatment is effective (Thompson et al do not deny this), just as they are part of the explanation why any medical treatment is effective[718]. The point however is that one cannot sustain, on this basis, the view that these components are the characteristic components of homeopathic treatment – they do not *characterise* it (the fact that homeopathic treatment might be particularly good at maximising the efficacy of its non-characteristic components will be discussed below).

If proponents of homeopathy intend to argue that the quasi-pharmacological component of homeopathic treatment is not the characteristic component (which is, of course, a legitimate strategy) then some account is owed of what makes some putative set of components characteristic of homeopathic treatment. In giving such an account, one would expect to be told how the dilution, dynamization, similarity and individualisation principles fit into the picture; since (as claimed in Part One) the use of these principles is a necessary condition of a treatment being homeopathic. On the most intuitive account, these principles are therapeutically relevant because they purportedly make a difference to the quasi-pharmacological contents of the

[718] (Edzard Ernst, 2011a; T. D. B. Thompson & Weiss, 2006; Zimmermann-Viehoff & Meissner, 2007)

medicines: the contents is potent due to the dilution and dynamisation principles, and the contents is applicable to the patient due to the similarity and individualisation principles. The justification in the Canonical Criticism for focusing on the quasi-pharmacological component of homeopathic treatment is that it is the most theoretically coherent candidate for being the characteristic component. This is the fifth conclusion:

Conclusion 5

If the quasi-pharmacological component of homeopathic treatment is not the characteristic component, then some account is owed of how any other candidate component could be characteristic of homeopathic treatment.

There is an important counter-argument to consider relating to Additivity, however. This is the second case to consider. A further analogy could be drawn, this time with the carisoprodol result discussed in Part Three. The legitimacy of a PCT would seem to depend on Additivity, as discussed in Part Three; in the case of the polypill, this assumption is warranted because it is known that the four component drugs of the polypill used in the PILL Collaboration study do not interact with each other and that their efficacy is robust across many circumstances[719]. In general however Additivity cannot be assumed without some evidence that it holds. The carisoprodol result provided a clear illustration of an instance in which it fails. Thus proponents of homeopathy may wish to make the argument that Additivity is false in the case of homeopathic treatment, and that this presents a barrier to discovering the efficacy of the quasi-pharmacological component of homeopathic treatment through placebo comparison. This is perhaps a better interpretation of the idea that homeopathic treatments are 'complex interventions', than is suggested by Thompson et al above. Indeed Weatherley-Jones et al made precisely this argument in Part One. To reiterate:

> 'The interaction of the non-specific effects of the consultation with the specific effects of the medicine appears to challenge

[719] (PILL Collaborative Group, 2011)

> the double-blind placebo-controlled RCT as a meaningful test of individualised[720] homeopathy[721],

I claim that this view is mistaken. Moreover I claim that the exact opposite is true. PCTs of individualised homeopathy represent the best case for investigating the efficacy of the quasi-pharmacological component of the treatment.

Consider again the carisoprodol result discussed in Part Three[722]. The efficacy of carisoprodol not only varied quantitatively but also changed qualitatively between different therapeutic contexts. When the symbolic dimension of the treatment was altered, so that patients were made to expect a stimulant effect from the carisoprodol pills they were receiving, the presence (compared to the absence) of carisoprodol augmented that effect. Carisoprodol (a relaxant, recall) stimulated the patients. When patients did not know what to expect, the pharmacological dimension of the treatment asserted itself and the presence of carisoprodol (compared to its absence) generated a relaxant effect. This illustrates a case where the result of a placebo comparison changed (in this case counter-intuitively) as a consequence of the context within which that comparison was set[723]. The reason for the difference in the carisoprodol case is that the expectations generated in one context were sufficient to modulate the patients' drug response.

Perhaps therefore proponents of homeopathy can be interpreted as making the claim that the efficacy of the quasi-pharmacological component of homeopathic treatment is modulated by the other components of homeopathic treatment; just as patients' expectations can modulate the effect of carisoprodol. Of course, it is an open question whether this is a good analogy to draw in the case of homeopathy. Evidence is needed for whether Additivity can, or cannot, be assumed in the case of homeopathy. As far as I know, this has not been investigated.

---

[720] As explained in Part One: individualised treatment is tailored to the patient, thus patients fitting into the same conventional disease category may not receive the same homeopathic treatment. In contrast with non-individualised homeopathy, where treatments are given in the matter of conventional drugs, so all patients with condition X receive pills containing homeopathic treatment Y.

[721] (Weatherley-Jones et al., 2004) p. 188

[722] (Flaten et al., 1999)

[723] There are plenty of further examples; another often cited study shows how the effect of medication for gastric ulcers differs according to cultural contexts. See: (Moerman, 2000)

I argue below that if proponents' arguments about complexity amount to the failure of Additivity due to some interactional effects between some of the different components of homeopathic treatment, then that does not entail that placebo comparison is a illegitimate way to investigate the efficacy of the quasi-pharmacological component of homeopathic treatment; however it does have methodological implications that must be taken into account to ensure the legitimacy of such comparisons.

Assume that the 'complexity' put forward by proponents of homeopathy amounts to a failure of Additivity. If the effect of the quasi-pharmacological component of homeopathy is modulated by other treatment components, then it is not clear why a placebo comparison should fail to find that is efficacious, *if that placebo comparison was set in the appropriate context*. That is, if those other treatment components were optimally in place. Weatherley-Jones et al, above, point to the mere fact of interaction between components as evidence that they cannot be investigated[724]. I claim that does not follow, however. Firstly one might investigate how to set-up the components of the treatment to maximise the efficacy of the quasi-pharmacological component; secondly one might investigate how robust the efficacy of the quasi-pharmacological component is to changes to those components (that is, changes in the therapeutic context).

The modulation of the efficacy of the quasi-pharmacological component by the other treatment components implies that the efficacy of the quasi-pharmacological component ought to be most demonstrable when experimental groups both receive the 'complete package' of homeopathic treatment. Setting-up treatment groups in circumstances that deviate substantially from typical homeopathic treatment are unlikely to reveal that the quasi-pharmacological component of homeopathy is efficacious, if (by a failure of Additivity) it's efficacy is highly sensitive to those other components. Whatever the nature of the interaction,

[724] They do note some more concrete issues in the design of trials of individualised homeopathy, such as the problem of homeopaths finding it difficult to assess the patients at follow-up, knowing they may be in the placebo group and thus being unsure how to interpret their patient's progress. This amounts to the claim that the efficacy of the quasi-pharamacological component is sensitive to such contextual changes; in which case some independent evidence is needed for that claim. The evidence must be independent because, if the failure of the trial to find a positive result is taken as evidence of such sensitivity, that would beg the question.

placebo-comparisons ought to be able to examine both the magnitude and sensitivity of the efficacy of the quasi-pharmacological component.

This provides a reason to design trials of homeopathic treatment which administer the homeopathic and control pills to each treatment group *in the context of the other components of homeopathic treatment*. It also provides a reason to question trials of homeopathic treatment that administer the homeopathic and control pills under circumstances that are atypical of normal homeopathic treatment. Importantly however, this is not an argument for the view that placebo comparison of the characteristic component of homeopathy is in principle impossible, even though Additivity may be false. Indeed if one takes seriously the fact that Additivity may be false for homeopathic treatment, then the 'in context' placebo comparison of homeopathy's characteristic component would seem to represent the best case for detecting an effect. This is the sixth conclusion:

Conclusion 6

Interaction between the different components of homeopathic treatment may present a legitimate problem when placebo controlled trials do not ensure both treatments groups also receive all those other non-characteristic components which are part of typical homeopathic treatment. In general trials should be designed to test and investigate Additivity. This is not a challenge to the, in principle, validity of placebo comparisons of homeopathic treatments, however.

13.2 The policy debate about homeopathy

In Part One the policy debate was described. Points (4) and (5), from §3.2.1 and the beginning of this chapter, were used to justify the claim that homeopathic treatment should not be available to patients. As the STC express it:

> 'to maintain patient trust, choice and safety, the Government should not endorse the use of placebo treatments, including homeopathy. Homeopathy should not be funded on the NHS and

> the MHRA [Medicines and Healthcare products Regulatory Agency] should stop licensing homeopathic products[725]'.

Points (4) and (5) both contend that 'placebo treatments' are unethical to provide. The STC's policy conclusions are reached by combining those points with the evidential claim that homeopathic treatment is a 'placebo treatment'. It has been explained how, on the basis of the arguments from Part Three, the notion of a 'placebo treatment' cannot be sustained. The most plausible reinterpretation of the notion is that by 'placebo treatments' the STC, and the Canonical Criticism more widely, mean that homeopathic treatments are effective treatments with an inefficacious characteristic component. What implications does this have for the policy debate?

Note that, of course, proponents of homeopathy contest the ethical arguments on evidential grounds. Proponents of homeopathy assert that that the characteristic component of homeopathic treatment is, in fact, efficacious. However, the following discussion will be conducted on the assumption that the characteristic component of homeopathic treatment is the quasi-pharmacological component and that it is indeed inefficacious. The purpose of this is to show that even on this assumption the conclusions which are drawn by the STC and in the Canonical Criticism about the impermissibility of providing homeopathic treatment do not follow from the ethical arguments they put forward.

It was argued in Part Three that treatments with an inefficacious characteristic component can be delivered in ways which do not necessary involve any deception of patients. However, it was also argued that the inefficacy of a treatment's characteristic component does introduce a tension between the therapeutic benefit from providing such a treatment and the potential deception that may result. Whilst not necessarily deceptive, treatments with inefficacious characteristic components seemed apt to result in deception unless special measures were taken to prevent patients being misled about the nature of their treatment. There are two questions to ask: first, how should the ethical arguments against the provision of homeopathic treatment be re-evaluated (that is, the No Placebos argument, and the Indirect Harm argument)? Second, are there circumstances under which the provision of homeopathic treatment might be permissible?

[725] (House of Commons Science & Technology Committee, 2010) para 157

Note firstly however that this discussion focuses on only some of the issues which the policy debate ranges over. The discussion will be concerned with whether a clinician can permissibly provide their patient with, or refer them for, homeopathic treatment. That is to say, I confine the discussion to the provision of homeopathic treatment by medically qualified practitioners – if homeopathic treatment is permissible in any setting, then it ought to be in that case. This also circumvents part of the 'Indirect Harm' argument made against homeopathic treatment. Specifically, it circumvents the worry that by receiving homeopathic treatment patients may delay the diagnosis of a more serious underlying condition, or simply forgo more effective conventional treatment. I take it that risk of harm through these means is minimised when patients are treated by medically qualified homeopaths. Or put another way, when homeopathic treatment is genuinely complementary, and not alternative, to conventional treatment. I will not examine this and other aspects of the Indirect Harm argument further, and I will not attempt to draw conclusions about the provision of homeopathy in other settings. Also I will not discuss the licensing or regulatory issues around the provision of homeopathic treatment.

Note secondly that the concern here is with the ethical arguments rather than policy recommendations. The arguments which purport to justify the policy position of the Canonical Criticism – and most particularly, the recommendations to the government put forward by the STC – are ethical. The ethical arguments are the more fundamental premises of the debate. The insights from Parts Two and Three clearly have implications throughout the policy debate (on regulation, health economics etc); however the discussion below attempts to draw out some of those implications only for the ethical arguments.

Note thirdly that the concern throughout this discussion is with the *permissibility* of providing homeopathic treatment. The concern is neither with whether there is an obligation to provide the treatment, nor with more practical questions of whether providing the treatment would be feasible, or cost-effective. The question is whether homeopathic treatment can be provided ethically: whether it is *worth* providing in certain circumstances or whether it *must* be provided in others are further separate questions. These will not be discussed.

13.2.1 The No Placebos argument

The No Placebos argument, described in Part One, cannot be sustained in the light of the arguments from Part Three. The No Placebos argument was based on the view that the effectiveness of placebo treatments is, necessarily, a product of false beliefs that patients have been deceived into holding. The deception involved was taken to be unethical for some of the traditional reasons given by bioethicists; for example, deception disregards the patients' autonomy and damages trust in the doctor-patient relationship[726].

That it cannot be sustained follows trivially from Part Three. Once the notion of 'placebos' and 'placebo effects' is abandoned, one cannot cite the fact that some treatment is a 'placebo' as evidence that it is unethical. This is not to deny that treatments which involve deceiving patients are ethically problematic. The point is that the reasons a treatment is or is not deceptive have nothing to do with whether it is called a 'placebo': because, as I have argued, there are no good reasons to call anything a 'placebo' besides arbitrary stipulation. The argument of Part Three demonstrates that it is not legitimate to argue homeopathic treatment is unethical *simply because* it is a 'placebo treatment' (as the No Placebos argument attempts to).

However, while I claim homeopathic treatment does not necessarily involve deception of patients in virtue of the fact that some call it a 'placebo', that claim leaves open the possibility that homeopathic treatment may involve deception in other ways. Importantly, Part Three argued that deception of patients was a more likely, but not a necessary, consequence of treatments with inefficacious characteristic components. Part Three also argued that treatments with inefficacious characteristic components still pose ethical problems, if the reasons for providing those treatments are not made explicit. The problem is that a treatment with an inefficacious characteristic component, if it is effective, is not effective (even partly) for the reason one would expect it to be.

Consequently, it may be possible for the basic conclusion of the No Placebos argument (namely: it is unethical to provide homeopathic treatments) to be recast in terms consistent with the argument of Part Three. Indeed, if one believes, as opponents of homeopathy do, that the characteristic, quasi-pharmacological,

[726] Specifically concerning the deceptive nature of homeopathic treatment, see: (House of Commons Science & Technology Committee, 2010) para 38 See also paras 94-101 (Baum & Edzard Ernst, 2009) p. 974 (Goldacre, 2007b) pp. 1672-3

component of homeopathic treatment is inefficacious, then a reinterpretation of the No Placebos argument, consistent with Part Three, seems like it ought to deliver similar conclusions.

The discussion about the ethical significance of placebo comparison in Chapter 11 suggests that there are at least two reasons why opponents of homeopathy could think that homeopathic treatment is ethically problematic. The first is directly related to the discussion in Part Three about the expectations that patients are warranted to form about the effectiveness of their treatments. The second is an argument that is available to opponents of homeopathy who hold a strong view about the mechanistic evidence, such as the view held in the STC report. After briefly outlining the two reasons that opponents of homeopathy could give, I shall go on to argue that this leads to a substantially weakened conclusion, in comparison to the original No Placebos argument described in Part One.

The first reason why homeopathic treatment may be ethically problematic concerns the use of pills. By providing patients with pills, patients would seem to be warranted in expecting that the effectiveness of the pills is partly (if not primarily) due to their pharmacological content. Again, assuming that the Canonical Criticism is correct with respect to the evidential debate, then given that the quasi-pharmacological content of homeopathic treatment is indeed inefficacious, giving patients pills is in danger of misleading them on this point. Consequently the justification for giving pills in homeopathic treatment cannot legitimately be that their quasi-pharmacological content is efficacious. Of course, giving pills is likely to be effective for other reasons[727], but the ethical tension stems from the fact that patients are justified in forming the false expectation that some or all of the effectiveness is due to the efficacy of the (quasi-)pharmacological content.

The second reason relies on the Implausibility Argument made in the Canonical Criticism. On the basis of the Implausibility Argument, the mechanistic evidence put forward by proponents of homeopathy is judged to be an inadequate basis from which to make inferences about putative therapeutic effects. Indeed, opponents of homeopathy believe that there can be no mechanism. Consequently opponents of homeopathy argue that the homeopathic principles of dilution, dynamization, similarity – in so far as they provide a mechanism by which the characteristic components of homeopathic treatment has its putative effects –

[727] See Chapter 8

provide no explanation of the efficacy of the characteristic component. Consequently reference to the proposed mechanism by which homeopathic treatment has it characteristic effects is straightforwardly deceptive, because it amounts to lying to patients. To justify the effectiveness of homeopathic treatment on the basis of an explanation that refers to homeopathic principles would be unethical because it would involve asserting more than can be justified[728].

These two reasons give some support to the view that homeopathic treatment is likely to involve some deception of patients, at least if they are given no special information about it. The fact that it involves providing pills which do not contain an efficacious pharmacological component, and the fact that the explanation for their effectiveness relies on a model that cannot work mean that providing patients with homeopathic pills and an explanation of effectiveness of the treatment which is consistent with homeopathic principles is likely to cause patients to be deceived about the nature of their treatment.

Note that, at best this is a weak argument for the view that it is unethical to provide homeopathic treatments, if only because it presupposes some un-evidenced empirical facts about why patients chose homeopathic treatment; for instance, that they do in fact care why it may work for them.

The problem for opponents of homeopathy is that it is not possible to hold onto the view that 'placebo treatments' are *necessarily* deceptive. Once reference to 'placebos' is removed, the question of whether a treatment involves deceiving patients is something to be assessed. It is true that the Canonical Criticism possesses the resources to motivate these two reasons just given, however these reasons are themselves rather weak, and also entirely defeatable if one can undertake measures to avoid any potential deception.

In Part Three is was suggested that in order to provide effective treatments with inefficacious characteristic components clinicians should be required to also provide explicit information to patients about the possible reasons why the treatment they received may be effective (This stood in contrast to other authors

[728] Note that this is not an argument against providing effective treatments when the explanation of the characteristic effects is not understood. It is an argument against providing effective treatments accompanied with an explanation of part of their effectiveness that is known to be unwarranted, such as when an explanation is known to be false (as is the case with homeopathy, according to the Canonical Criticism).

who argued for less explicit 'information conditions' on the provision of such treatments [729]). In what follows I will consider whether there are, in fact, circumstances under which the provision of homeopathic treatment may be permissible. I claim that the problem posed by the two potentially deceptive elements identified above can, largely, be circumvented. Firstly however, note the seventh conclusion to draw:

Conclusion 7

The No Placebos argument fails. The provision of homeopathic treatment is not necessarily *deceptive. Consequently there may be circumstances under which it is permissible to provide it.*

13.2.2 A possible role for homeopathic treatment

There are two possible roles for homeopathic treatment that I wish to consider, the first and least controversial concerns the utility of adopting homeopathic consulting practices in conventional treatment; that is, a role for homeopathic treatment as a possible resource for conventional medicine to learn from. The second is more controversial. This concerns the possible circumstances in which homeopathic treatment, as such, might be provided to patients even if the Canonical Criticism is correct that the quasi-pharmacological component of homeopathic treatment is inefficacious.

The first and least controversial way to remove the ethical problems with homeopathic treatment is to export the effective non-characteristic components to other ethically acceptable medical treatments. Homeopathic treatment is made up of a configuration of non-characteristic components that could plausibly be seen as suggesting ways to modify other treatments, which share some or all of those components. Of course some components, such as the patients' belief *in homeopathic treatment* will not export.

As Part Three argued, there are acceptable ways to maximise therapeutic effects of the non-characteristic components of treatments that do not involve deceiving patients. A number of authors have pointed out the non-characteristic

[729] See: (Lichtenberg et al., 2004; Pittrof & Rubenstein, 2008)

components of homeopathy, such as the long consultation, are indeed likely to be efficacious in their own right[730]. That is to say, homeopathic treatment seems to provide an exemplary configuration of non-characteristic components that improve therapeutic benefit. The key question is whether the way these components are utilised in homeopathic treatment can be carried over into conventional treatments.

This suggestion is speculative: it involves extrapolating from theoretical points. Never the less testable hypotheses follow from taking the idea seriously. For example, If GP consultations followed a more homeopathic model, would this have significant therapeutic consequences? How would such changes to conventional practice compare to homeopathic treatment of the same conditions?[731] How do different components of homeopathic treatment interact to create greater overall treatment effectiveness? If one is interested in improving the effectiveness of treatments, through maximising the efficacy of non-characteristic components of treatment, then I claim that homeopathic treatment provides an excellent case study for empirical investigations. In so far as homeopathic treatment consists of a consultation followed by prescription of pills, it provides a good model of many conventional treatment contexts; unlike other alternative treatments (for example, acupuncture). Or to put this another way, I suggest that investigating homeopathic treatment could be part of a research agenda, along the lines currently pursued in somewhat artificial circumstances by, for example, the research into 'placebo phenomena' (Chapter 9) but one which is, potentially, a more clinically relevant research agenda. This is the eighth conclusion:

Conclusion 8

One possible role for homeopathic treatment is as a subject for research. It provides a (perhaps more clinically relevant) alternative to the situations typically studied by researchers investigating 'placebo effects'. For example, how effective are the non-characteristic components of homeopathic treatment, in practice?

[730] See: (T. D. B. Thompson & Weiss, 2006) See also: (Edzard Ernst, 2011a; Zimmermann-Viehoff & Meissner, 2007)

[731] Some evidence is available for the additional benefit provided by augmenting conventional treatment with homeopathic treatment. For example, Relton et al found a positive effect from conventional plus homeopathic treatment of fibromyalgia. See: (Relton et al., 2009)

The second role for homeopathic treatment is as an ethically acceptable treatment in its own right. Consider from above that, according to the Canonical Criticism, the problem with homeopathic treatment is that its characteristic component is inefficacious. How much and what kind of an ethical problem is this however? - I have argued that this makes it more likely that patients will be deceived about why the treatment is effective, but that no deception is necessary. I suggest that given this, there are circumstances in which it is permissible to provide homeopathic treatment. To reiterate: I suggest that there are circumstances in which it is permissible to provide homeopathic treatment, *on the evidential assumptions of the Canonical Criticism*. Following the criticism of Pittrof and Rubenstein[732], and also Lichtenberg's [733] criteria for the permissible provision of 'placebo treatments' discussed in Chapter 10, I suggest that a clinician may permissibly provide homeopathic treatment only if[734]:

Effectiveness condition:

(1) there is either no, or no more effective, conventional treatment;

OR

(2) the patient explicitly wants homeopathic treatment and they are aware of any substantially more effective conventional treatments;

AND[735]

Information condition:

(3) The patient is aware that very good evidence suggests that the effectiveness of homeopathic treatment is unlikely to be due to the quasi-pharmacological content of the pills, and unlikely to be explained by homeopathic principles[736].

[732] (Pittrof & Rubenstein, 2008)

[733] (Lichtenberg et al., 2004)

[734] I do not claim these conditions are jointly sufficient, merely jointly necessary (hence 'only if' not 'if' or 'iff'). It may be the case that provision of homeopathy treatments is unacceptable for other reasons, perhaps concerning the Indirect Harm argument discussion of which has been put aside here.

[735] The indent of the connectives indicates their scope, i.e. (1 v 2) & 3 – of course 1 OR 2 does not rule out 1 AND 2; that is, one could fulfil both.

[736] That is to say, the patient is told honestly that the effectiveness of the treatment is explained by the efficacy of the treatment's non-characteristic components – for example, that they will be responding to the symbolic components of the treatment; such as the attentive consultation, the fact they *receive* pills and not the *contents* of the pills etc.

Notice first that (1)-(3) would clearly be unacceptable to proponents of homeopathy, if only because of the assumption that the characteristic component is inefficacious[737]. Notice second that the justification for the view that homeopathic treatment can permissibly be provided only if (1)-(3) rests on the idea that the potential deception involved in homeopathic treatment can be mitigated by the circumstances under which it is provided. To what extent do (1)-(3) succeed in this?

(1) & (2) attempt to ensure that patients are not misled about the availability of other conventional treatments. (1) allows the provision of homeopathic treatment if there are no better alternatives. (2) allows the provision of homeopathic treatment, in spite of better alternatives, if the patient's own values and preferences are such that they strongly want homeopathic treatment. (1) & (2) attempt to ensure that when patients do exercise their autonomy, they are not basing their choices on false premises. Homeopathic treatment should not be a treatment option when there are more effective treatments, unless the patient has a strong preference for it and they understand what they may be giving up. As with any medical treatment, it remains the patient's choice whether or not to accept homeopathic treatment, even if there is no more effective alternative.

In order that patients make an informed choice, (3) stipulates a further necessary condition. (3) is similar to what were termed 'information conditions' in the discussion of Pittrof and Rubenstein, and Lichtenberg in Chapter 10; however it is more stringent since it requires that the information should be given without the patient having to ask the right questions. (3) Requires that patients should be aware of why homeopathic treatment is likely to be effective. Since this argument is being put forward on the same assumptions made in the Canonical Criticism, this means that patients should understand that there is very good evidence that the effectiveness of the treatment is due only to the non-characteristic components of the treatment. That is to say, patients should understand that the quasi-pharmacological component of the treatment is inefficacious. Moreover they ought to understand that they are being given a pill on account of the efficacy of the

[737] Consider for example this statement, from a letter in the journal *Trends in Pharmaceutical Sciences*: '[on the question of] whether it is ethical for homeopaths to use a placebo if they know it is only a placebo. This debate is irrelevant; homeopaths know they are providing more than a placebo' (Ross, 2010) p. 297

symbolism and medico-cultural associations that pill-giving generates, not because of the contents of the pill.

If patients have the awareness required by (3), then patients ought to have (according to the assumptions of the Canonical Criticism) an evidence-based view about the effectiveness of homeopathic treatment. In this way, (3) attempts to respect a patients autonomy and choice: it seeks to ensure that patients are not offered treatments under false pretences or on bad faith. (3) ought to prevent patients being given true, but misleading, information about the effectiveness of homeopathic treatment. The aim is that the patient should be aware of the evidence-base for why the treatment is effective and the clinician's true rationale for providing it.

There may seem to be something absurd about (3). It might be argued that, if one insists patients are told why homeopathic treatment is effective, then patients would refuse it. This would be an interesting empirical claim to test. Identification with the philosophical and 'natural' principles that underlie many alternative medicines has been shown to be a key driver of patients' use of them[738]. On the other hand the results reported by Kaptchuk et al are also a relevant point to consider[739]. Katpchuk et al conducted a randomised trial comparing open-label 'placebo pills[740]' to no-treatment for treatment of irritable bowel syndrome. They showed that even when patients were told they were receiving a 'placebo pill' and had the likely reasons for that pill's effectiveness explained to them, patients were still happy to take the pill and experienced significant improvements on the main and secondary outcome measures. I simply conclude that in the absence of research investigating the question, it is simply not obvious how patients would respond to being offered homeopathic treatment under circumstances constrained by (1)-(3).

The scenario that the STC and the Canonical Criticism invite one to imagine in their ethical arguments consists of a patient being given homeopathic treatment

[738] (Bishop, Yardley, & Lewith, 2007, 2010; Furnham, 1996)
[739] (Kaptchuk et al., 2010)
[740] More precisely: prior to randomisation all patients had the 'placebo effect' explained to them so they understood why it was the pills they may be given were likely to have an effect (this was validated at follow-up by a semi-structured survey, which showed that patients generally understood the nature of the 'placebo' treatment). Those assigned to the treatment group were given 'a typical prescription medicine bottle of placebo pills with a label clearly marked ''placebo pills'' ''take 2 pills twice daily.'' The placebo pills were blue and maroon gelatin capsules filled with avicel, a common inert excipient for pharmaceuticals' (Kaptchuk et al., 2010) pp. 2-3

accompanied by claims that the pills have an efficacious pharmacological content, and that the treatment is effective because of facts about the patient's symptoms and their relation to homeopathic principles[741]. I agree that, on the view about the evidential debate held by the Canonical Criticism, providing homeopathic treatment on these terms would be unethical. However, I also claim that there are alternative scenarios where, on the same evidential assumptions, it is permissible to provide homeopathic treatment. As (1)-(3) set out, providing homeopathic treatment is permissible when the patient both understands that the effectiveness of the treatment is due solely to its non-characteristic components and prefers homeopathic treatment over any conventional alternatives. Stripped of reference to 'placebos', a more general statement of the problem that (1)-(3) aim to address is: how can clinicians provide effective treatments ethically? Homeopathic treatment is a difficult case because, unusually, it is effective but its characteristic component is inefficacious. My claim, in suggesting (1)-(3), is that the principle 'provide effective treatments ethically' need not prohibit the provision of homeopathic treatment.

The most obvious objection to this is that it seems to imply that the widespread use of any treatment with an inefficacious characteristic component would be permissible in analogous circumstances. The problem is that (1)-(3) generalises in such a way that it would be permissible to provide *any* exotic or fanciful treatment, if it were effective and if it were accompanied with appropriate information about why it was effective.

Consider again the example from Part Three of treating a headache with head-patting. It seems unlikely that the physical-patting component of the treatment would be efficacious. Never the less, let us assume that such patting could be delivered in a context in which the patting was performed by a physician and that physical contact of this sort was deeply symbolic – such that these symbolic components of the treatment were truly efficacious. Further assume that, for this reason, the treatment is actually effective for headaches. We could generate, with suitable modifications, conditions (1*)-(3*) analogous to (1)-(3) which specified conditions for permissibly providing head-patting for headaches. More generally and more importantly, relevant modifications of (1)-(3) could be produced for any kind of

[741] A nice illustration of the scenario that the STC use to frame the debate can be seen in Q205 of the second oral evidence session (House of Commons Science & Technology Committee, 2010) Ev. 68

exotic and fanciful, *but effective*, treatment that one might care to invent. The objection is that this is an unacceptable generalisation.

The generalisation is valid. I deny however that it is unacceptable. Firstly, in cases where there are more effective conventional alternatives to these exotic or fanciful treatments, the question of whether it is permissible to provide the treatment depends on the patient's values and preferences in the face of medical knowledge. Secondly, in cases where there are no more effective conventional treatments, it is difficult to see what the objection could be, if the patient is aware of the nature of the alternative treatment's effectiveness. Of course there may be reasons not to provide a treatment, based on facts about its cost-effectiveness, or the low demand for it (in short practical reasons). Crucially however it is not those kinds of reasons that opponents of homeopathy put forward; they rely on the more fundamental ethical objection to 'placebo treatments'. I claim that the ethical objection fails in the simple case where a fully informed patient expresses the desire for such a treatment. This gives a tentative and final ninth conclusion:

Conclusion 9

A patient's fully informed desire for a demonstrably less effective treatment is sufficient to make the provision of that treatment permissible[742]*. Even on the evidential assumptions of the Canonical Criticism, it is permissible to provide homeopathic treatment under appropriate circumstances outlined in §13.2.2.*

[742] This ought not to be controversial – since it seems only to embody the notion, central to the EBM view, that best evidence must be integrated with patient values and circumstances.

CHAPTER 14

14. Conclusion

The evidential and policy debates in the homeopathy controversy draw on the concepts of evidence-based medicine and placebos in many of the arguments put forward by both proponents and opponents of homeopathy.

There is no single or stable interpretation of EBM in the medical literature. The literature is, and always has been, unclear about what the details of EBM amount to. The most straightforward interpretation, the Categorical Interpretation, is not defensible. There are no a priori constraints on what kinds of methods can generate good evidence. Importantly therefore, any evaluation of the evidence for the efficacy of the quasi-pharmacological component of homeopathic treatment ought to take into account the mechanistic evidence for and against its purported efficacy. In this respect the House of Common's Science & Technology Committee undervalue mechanistic evidence. This is because they hold a Categorical Interpretation of EBM. This is of particular significance because, in fact, they endorse strong claims about the mechanistic evidence against the efficacy of the quasi-pharmacological component of homeopathic treatment.

When it comes to evaluating whether homeopathic treatment 'works', the key concern is with the efficacy of the characteristic component. The efficacy of the characteristic component is only important for ethical reasons. Opponents of homeopathy who claim it does not 'work' must be seen as expressing an ethical objection to the reasons why it is effective. It is plausible to assume that the quasi-pharmacological component of homeopathic treatment is the characteristic component. If the quasi-pharmacological component of homeopathic treatment is not the characteristic component, then proponents of homeopathy owe some account of how any other candidate component could be characteristic of homeopathic treatment. Complex interactions between different treatment components might be relevant to consider here, and investigating the validity of Additivity in relation to homeopathic treatment should be an important aspect of trail design. Notice however that while interaction between the different components of homeopathic treatment may present a legitimate problem when placebo controlled trials do not ensure both experiment groups also receive all those

other non-characteristic components which are part of typical homeopathic treatment, this is not a challenge to the, in principle, validity of placebo comparisons of homeopathic treatments.

The argument that it is unethical to provide homeopathy because of the deceptive nature of 'placebos' fails. The provision of homeopathic treatment is not necessarily deceptive. Consequently there may be circumstances under which it is permissible to provide it. One possible role for homeopathic treatment is as a subject of research, since it provides a (perhaps more clinically relevant) alternative to the situations typically studied by researchers investigating 'placebo effects'. A second possible role is suggested by the idea that a patient's fully informed desire for a demonstrably less effective treatment is sufficient to make the provision of that treatment permissible. Even on the evidential assumptions of the Canonical Criticism, it may be permissible in certain circumstances to provide homeopathic treatment when there is no better alternative, the patient wants homeopathic treatment, and the patient is aware of the reasons why (according to the Canonical Criticism) homeopathic treatment is effective.

REFERENCES

10:23 Campaign. (2010). Homeopathy: there's nothing in it. Retrieved December 2010, from http://www.1023.org.uk/

Adler, H. M., & Hammett, V. (1973). The doctor-patient relationship revisited: An analysis of the placebo effect. Annals of Internal Medicine, 78(4), 595-598.

Adler, S. R. (2011). Sleep Paralysis: Night-mares, Nocebos, and the Mind-Body Connection. London: Rutgers University Press.

Adolphs, S. (2006). Introducing electronic text analysis. London: Routledge.

Alexa, M. (1997). Computer-assisted text analysis methodology in the social sciences. Retrieved February 2010, from <http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/zuma_arbeitsberichte/97_07.pdf>

Altunç, U., Pittler, M. H., & Ernst, Edzard. (2007). Homeopathy for Childhood and Adolescence Ailments: Systematic Review of Randomized Clinical Trials. Mayo Clinic proceedings, 82(1), 69-75.

Amanzio, M., & Benedetti, F. (1999). Neuropharmacological dissection of placebo analgesia: expectation-activated opioid systems versus conditioning-activated specific subsystems. Journal of Neuroscience, 19(1), 484-494.

Amanzio, M., Pollo, A., Maggi, G., & Benedetti, F. (2001). Response variability to analgesics: a role for non-specific activation of endogenous opioids. Pain, 90(3), 205-215.

Anekwe, L. (2010). Ten PCTs pull NHS funding for homeopathy. Pulse News. Retrieved August 2010, from http://www.pulsetoday.co.uk/story.asp?sectioncode=23&storycode=4126832

Archie Cochrane Archive Catalogue: ALC/5 Health Services Research. Retrieved November 2008, from http://www.cardiff.ac.uk/insrv/libraries/scolar/archives/cochrane/warexperience.html

Aronson, J. K., & Hauben, M. (2006). Drug safety Anecdotes that provide definitive evidence. British Medical Journal, 333, 1267-1269.

Ashcroft, R. E. (2004). Current epistemological problems in evidence based medicine. Journal of Medical Ethics, 30(2), 131-135.

Asthana, A., & McKie, R. (2010). Homeopathic remedies : a real cure or a waste of NHS money ? Guardian. Retrieved January 31, 2010, from http://www.guardian.co.uk/lifeandstyle/2010/jan/31/homeopathic-remedies-nhs

BBC Radio 4. (2005). Case Notes: Homeopathy. BBC. Retrieved July 19, 2005, from http://www.bbc.co.uk/radio4/science/casenotes_tr_20050719.shtml

Baker, M., & Davenport, A. (2011). Homeopathy and other quackery at the Science Museum. Guardian. Retrieved February 9, 2011, from http://www.guardian.co.uk/science/the-lay-scientist/2011/feb/08/1

Ball, P. (2008). Water: water--an enduring mystery. Nature, 452(7185), 291-2.

Barratt, A., Irwig, L., Glasziou, P., Cumming, R. G., Raffle, A., Hicks, N., Gray, J. A. M., et al. (1999). Users' Guides to the Medical Literature: XVII. How to Use Guidelines and Recommendations About Screening. JAMA, 281(21), 2029-2034.

Barrett, B., Muller, D., Rakel, D., Rabago, D., Marchand, L., & Scheder, J. (2006). Placebo, meaning, and health. Perspectives in Biology and Medicine, 49(2), 178-198.

Barry, C. A. (2006). The role of evidence in alternative medicine: Contrasting biomedical and anthropological approaches. Social Science & Medicine, 62, 2646-2657.

Barsky, A. J., Saintfort, R., Rogers, M. P., & Borus, J. F. (2002). Nonspecific medication side effects and the nocebo phenomenon. JAMA, 287, 622-627.

Baum, M. (2004). An open letter to the Prince of Wales: with respect, your highness, you've got it wrong. British Medical Journal, 329, 118.

Baum, M. (2006). Re Use of " alternative " medicine in the NHS. The Times. Retrieved May 23, 2006, from http://www.timesonline.co.uk/tol/news/uk/health/article723787.ece

Baum, M., & Ernst, Edzard. (2009). Should We Maintain an Open Mind about Homeopathy? American Journal of Medicine, 122(11), 973-974.

Beauchamp, T. L., & Childress, J. F. (2008). Principles of Biomedical Ethics (Sixth Edition). Oxford: Oxford University Press.

Beecher, H. K. (1955). The powerful placebo. JAMA, 159(17), 1602-1606.

Bell, I. R. (2005). All Evidence Is Equal, but Some Evidence Is More Equal than Others: Can Logic Prevail over Emotion in the Homeopathy Debate? Journal of Alternative & Complementary Medicine, 11(5), 763-769.

Bell, I. R., & Koithan, M. (2006). Models for the Study of Whole Systems. Integrative Cancer Therapies, 5(4), 293-307.

Benedetti, F. (2009). Placebo Effects: Understanding the mechanisms in health and disease. Oxford: Oxford University Press.

Benedetti, F., & Amanzio, M. (1997). The neurobiology of placebo analgesia: from endogenous opioids to cholecystokinin. Progressive Neurobiology, 52(2), 109-125.

Benedetti, F., Amanzio, M., & Maggi, G. (1995). Potentiation of placebo analgesia by proglumide. Lancet, 346, 1231.

Benedetti, F., Pollo, A., Lopiano, L., Lanotte, M., Vighetti, S., & Rainero, I. (2003). Conscious expectation and unconscious conditioning in analgesic, motor, and hormonal placebo/nocebo responses. Journal of Neuroscience, 23(10), 4315-4323.

Berger, J. (2009). Paternalistic assumptions and a purported duty to deceive. American Journal of Bioethics, 9(12), 20-21.

Berghmans, R., & Schouten, H. C. (2011). Sir Karl Popper, swans, and the general practitioner. British Medical Journal, 343, d5469.

Bird, A. (2005). Abductive Knowledge and Holmesian Inference. Oxford Studies in Epistemology, (Vol. 1). Oxford: Oxford University Press.

Bird, A. (2011). The epistemological function of Hill's criteria. Preventive Medicine, 53(4-5), 242-245.

Bishop, F. L., Yardley, L., & Lewith, G. T. (2007). A Systematic Review of Beliefs Involved in the Use of Complementary and Alternative Medicine. Journal of Health Psychology, 12(6), 851-867.

Bishop, F. L., Yardley, L., & Lewith, G. T. (2010). Why Consumers Maintain Complementary and Alternative Medicine Use: A Qualitative Study. Journal of Alternative & Complementary Medicine, 16(2), 175-182.

Bivins, R. (2007). Alternative medicine? A History. Oxford: Oxford University Press.

Blackie, M. (1981). The Challenge of Homeopathy. London: Unwin.

Blackie, M. (1986). Classical Homeopathy. Beaconsfield: Beaconsfield Publishers.

Blackwell, B., Bloomfield, S. S., & Buncher, C. R. (1972). Demonstration to medical students of placebo responses and non-drug factors. Lancet, 1, 1279-1282.

Blasi, Z. D., Harkness, E., Ernst, Edzard, Georgiou, A., & Kleijnen, J. (2001). Influence of context effects on health outcomes: a systematic review. Lancet, 357, 757-762.

Bluhm, R. (2005). From hierarchy to network: a richer view of evidence for evidence-based medicine. Perspectives in Biology and Medicine, 48(4), 535-547.

Bluhm, R. (2010). Evidence-based medicine and philosophy of science. Journal of Evaluation in Clinical Practice, 16(2), 363-364.

Boer, A. D., & Porsius, A. J. (1997). Why do scientists still study and discuss homeopathy? Pharmacy World and Science, 19(4), 176-177.

Boiron, C. (2011). Homeopathy, a tremendous opportunity for medicine! European Journal of Internal Medicine, 22(1), 117-118.

Boissel, J.-P., Cucherat, M., Haugh, M., & Gauthier, E. (1996). Critical literature review on the effectiveness of homoeopathy: overview of data from homoeopathic medicine trials. Homoeopathic Medicine Research Group. Report to the European Commission: Brussels.

Borg, I., & Groenen, P. (1997). Modern Multidimensional Scaling. New York: Springer.

Borgerson, K. (2009). Valuing Evidence: bias and the evidence hierarchy of evidence-based medicine. Perspectives in Biology and Medicine, 52(2), 218-233.

Borland, D. (1988). Homeopathy in Practice. Beaconsfield: Beaconsfield Press.

Boseley, S. (2010). Stop funding homeopathy, MPs urge. Guardian. Retrieved February 22, 2010, from http://www.guardian.co.uk/society/2010/feb/22/stop-funding-homeopathy-mps-urge

Branthwaite, A., & Cooper, P. (1981). Analgesic effects of branding in treatment of headaches. British Medical Journal, 282, 1576-1578.

British Homeopathic Association. (2010). comments on the recommendations of the Science & Technology Committee's "Evidence Check 2: Homeopathy."

Brody, H. (1982). The Lie That Heals: The Ethics of Giving Placebos. Annals of Internal Medicine, 97, 112-118.

Brody, H. (2009). Medicine's continuing quest for an excuse to avoid relationships with patients. American Journal of Bioethics, 9(12), 13-15.

Brody, H., Miller, F. G., & Bogdan-Lovis, E. (2005). Evidence-based medicine: watching out for its friends. Perspectives in Biology and Medicine, 48(4), 570-84.

Brooks, M. (2009). Homeopathy: Sometimes a dose of nothing can do you a power of good. Guardian. Retrieved February 6, 2009, from http://www.guardian.co.uk/science/blog/2009/feb/06/homeopathy-homeopathic-nhs-placebo-effect

Bucher, H. C., Guyatt, G. H., Cook, D. J., Holbrook, A., & McAlister, F. A. (1999). Users' Guides to the Medical Literature: XIX. Applying Clinical Trial Results; A. How to Use an Article Measuring the Effect of an Intervention on Surrogate End Points. JAMA, 282(8), 771-778.

Buetow, S. (2009). EBM and the strawman: a commentary on Devisch and Murray (2009). "We hold these truths to be self-evident": deconstructing 'evidence-based' medical practice. Journal of Evaluation in Clinical Practice, 15, 957-959.

Buetow, S., Upshur, R., Miles, A., & Loughlin, M. (2006). Taking stock of evidence-based medicine: opportunities for its continuing evolution. Journal of Evaluation in Clinical Practice, 12(4), 399-404.

Butterworth, S. (2007). Open door: The Guardian reader's editor on ... the homeopathy debate. Guardian. Retrieved August 2009, from http://www.guardian.co.uk/commentisfree/2007/dec/10/comment.leadersandreply?INTCMP=SRCH

Campbell, A. (1984). The Two Faces of Homeopathy. London: Hale.

Campbell, M., Fitzpatrick, R., Haines, A., Kinmonth, A. L., Sandercock, P., Spiegelhalter, D., & Tyrer, P. (2000). Framework for design and evaluation of complex interventions to improve health. British Medical Journal, 321, 694-696.

Campbell, N. C., Murray, E., Darbyshire, J., Emery, J., Farmer, A., Griffiths, F., Guthrie, B., et al. (2007). Designing and evaluating complex interventions to improve health care. British Medical Journal, 334, 455-459.

Canadian Task Force On The Periodic Health Examination. (1979). The periodic health examination. Canadian Medical Association Journal, 121, 1193-1254.

Carnap, R. (1950). Logical Foundations of Probability. Chicago: University of Chicago Press.

Cartwright, N. (2007). Are RCTs the Gold Standard? BioSocieties, 2(1), 11-20.

Cartwright, N. (2009). What are randomised controlled trials good for? Philosophical Studies, 147(1), 59-70.

Cartwright, N. (2011a). A philosopher's view of the long road from RCTs to effectiveness. Lancet, 377, 1400-1401.

Cartwright, N. (2011b). Predicting what will happen when we act. What counts for warrant ? Preventive Medicine, 53(4-5), 221-224.

Cartwright, N. (2009) What is This Thing Called "Efficacy"? In C. Mantzavinos, ed. *Philosophy of the Social Sciences*. Cambridge: Cambridge University Press.

Cartwright, N., & Munro, E. (2010). The limitations of randomized controlled trials in predicting effectiveness. Journal of Evaluation in Clinical Practice, 16(2), 260-266.

Caspi, O., & Bootzin, R. R. (2002). Evaluating How Placebos Produce Change: Logical and Causal Traps and Understanding Cognitive Explanatory Mechanisms. Evaluation & the Health Professions, 25(4), 436-464.

Cattaneo, A. D., Lucchelli, P. E., & Filippucci, G. (1970). Sedative effects of placebo treatment. European Journal of Clinical Pharmacology, 3(1), 43-45.

La Caze, A. (2008). Evidence-Based Medicine Can't Be…. Social Epistemology, 22(4), 353-379.

La Caze, A. (2009). Evidence-Based Medicine Must Be …. Journal of Medicine and Philosophy, 34(5), 509-527.

La Caze, A. (2011). The role of basic science in evidence-based medicine. Biology & Philosophy, 26(1), 81-98.

La Caze, A., Djulbegovic, B., & Senn, S. (2011). What does randomisation achieve? [EARLY VIEW]. Evidence Based Medicine, 1-2. doi:10.1136/ebm.2011.100061

Chalmers, I., Sackett, D. L., Silagy, C., & Maynard, A. (1997). The Cochrane Collaboration. London: BMJ Publishing Group.

Chaplin, M. F. (2007). The Memory of Water: an overview. Homeopathy, 96(3), 143-150.

Charlton, B. G., & Miles, A. (1998). The rise and fall of EBM. QJM, 91(5), 371-374.

Chatfield, K. (2008). In Pursuit of Evidence. Society of Homeopaths. Retrieved January 7, 2008, from http://www.homeopathy-soh.org/whats-new/researchwn.aspx

Cheatham, M. L. (2008). The death of George Washington: an end to the controversy? American Surgeon, 74(8), 770-774.

Cherkin, D. C., Sherman, K. J., Avins, A. L., Erro, J. H., Ichikawa, L., Barlow, W. E., Delaney, K., et al. (2009). A Randomised Trial Comparing Acupuncture, Simulated Acupuncture, and Usual Care for Chronic Low Back Pain. Archives of Internal Medicine, 169(9), 858-866.

Clover, A. (1989). Homeopathy Reconsidered. London: Victor Gollancz.

Cobb, L. A., Thomas, G. I., Dillard, D. H., Merendino, K. A., & Bruce, R. A. (1959). An evaluation of internal-mammary-artery ligation by a double-blind technic. New England Journal of Medicine, 260(22), 1115-1118.

Cochrane Collaboration. (2009). The Cochrane Logo. Retrieved March 9, 2009, from http://www.cochrane.org/logo/

Cochrane, A. (1945). Tuberculosis among Prisoners of War in Germany. British Medical Journal, 2(4427), 656-658.

Cochrane, A., & Blythe, M. (1989). One Man's Medicine: An Autobiography of Professor Archie Cochrane. London: British Medical Journal Books.

Cohen, A. M., Stavri, P. Z., & Hersh, W. R. (2004). A categorization and analysis of the criticisms of Evidence-Based Medicine. International Journal of Medical Informatics, 73(1), 35-43.

Cohen, B. (2005). The death of George Washington (1732-99) and the history of cynanche. Journal of Medical Biography, 13(4), 225-231.

Cohen, Deborah. (2010). BMA meeting: BMA representatives vote to ban homoeopathy from the NHS. British Medical Journal, 340, c3513.

Cohen, D. (2011). Edzard Ernst: the prince and me. British Medical Journal, 343, d4937.

Cohen, Deborah. (2009). Agency is criticised after granting a licence for a homeopathic remedy. British Medical Journal, 338, 1229-1230.

Cohen, N. (2007). The cranks who swear by citronella oil. Guardian. Retrieved October 28, 2007, from http://www.guardian.co.uk/commentisfree/2007/oct/28/comment.health?INTCMP=SRCH

Colloca, L, & Benedetti, F. (2005). Placebos and painkillers: Is mind as real as matter. Nature Reviews Neuroscience, 6(7), 545-552.

Colloca, Luana, & Miller, F. G. (2011). How placebo responses are formed: a learning perspective. Philosophical Transactions of the Royal Society B, 366(1572), 1859-1869.

Colquhoun, D. (2009a). Secret remedies: 100 years on. British Medical Journal, 339, b5432.

Colquhoun, D. (2009b). MHRA label seems to be illegal. British Medical Journal, 338, 1404-1405.

Colquhoun, D. (2011). DC's Improbable Science blog. <http://www.dcscience.net/>

Comaroff, J. (1976). A bitter pill to swallow: placebo therapy in general practice. Sociological Review, 24(1), 79-96.

Concato, J. (2004). Observational versus experimental studies: what's the evidence for a hierarchy? NeuroRx, 1(3), 341-347.

Concato, J., Shah, N., & Horwitz, R. I. (2000). Randomized, controlled trials, observational studies, and the hierarchy of research designs. New England Journal of Medicine, 342(25), 1887-1892.

Conrad, P. (1992). Medicalization and social control. Annual Review of Sociology, 18, 209–232.

Coulter, H. L. (1975). Divided legacy: A history of the schism in medical thought (Vol. II). Berkley, CA: North Atlantic Books.

Cox, T. F., & Cox, M. A. A. (2001). Multidimensional Scaling (Second.). London: Chapman & Hall/CRC.

Coxon, A. P. M., & Davies, P. M. (1982). The User's Guide to Multidimensional Scaling. London: Heinemann Educational Books.

de Craen, A. J. M., Moerman, D. E., Heisterkamp, S. H., Tytgat, G. N. J., Tijssen, J. G. P., & Kleijnen, J. (1999). Placebo effect in the treatment of duodenal ulcer. British Journal of Clinical Pharmacology, 48(6), 853-860.

de Craen, A. J. M., Roos, P. J., De Vries, A. L., & Kleijnen, J. (1996). Effect of colour of drugs: systematic review of perceived effect of drugs and of their effectiveness. British Medical Journal, 313(7072), 1624-1626.

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2008). Developing and evaluating complex interventions: the new Medical Research Council guidance. British Medical Journal, 337, 979-983.

Craig, Peter, Dieppe, Paul, Macintyre, Sally, Michie, Susan, Nazareth, Irwin, & Petticrew, Mark. (2008). Developing and evaluating complex intervention: new guidance. London: Medical Research Council.
<http://www.mrc.ac.uk/Utilities/Documentrecord/index.htm?d=MRC004871>

Crombie, I. K. (2008). The Pocket Guide to Critical Appraisal. London: BMJ Publishing Group.

Csordas, T. J. (1988). Elements of Charismatic Persuasion and Healing. Medical Anthropology Quarterly, 2(2), 121-142. Blackwell Publishing on behalf of the American Anthropological Association.

Cucherat, M., Haugh, Margaret, Gooch, M., & Boissel, J.-P. (2000). Evidence of clinical efficacy of homeopathy: A meta-analysis of clinical trials. European Journal of Clinical Pharmacology, 56, 27-33.

Daly, J. (2005). Evidence-Based Medicine and the Search for a Science of Clinical Care. London: University of California Press.

Danciger, E. (1987). The Emergence of Homeopathy. London: Century.

Dans, A. L., Dans, L. F., Guyatt, G. H., & Richardson, W. S. (1998). Users' Guides to the Medical Literature: XIV. How to Decide on the Applicability of Clinical Trial Results to Your Patient. JAMA, 279(7), 545-549.

Dantas, F. (2005). Are the clinical effects of homoeopathy placebo effects? Lancet, 366, 2083.

Davidovitch, Nadav, & Filc, D. (2006). Reconstructing data: evidence-based medicine and evidence-based public health in context. Dynamis, 26, 287-306.

Davies, P. M., & Coxon, A. P. M. (1982). Key Texts in Multidimensional Scaling. London: Heinemann Educational Books.

Department Of Health. (2010). Government Response to the Science and Technology Committee report 'Evidence Check 2: Homeopathy'. London. <http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_117810>

Department of Clinical Epidemiology & Biostatistics McMaster University. (1981a). How To Read Clinical Journals: II. To learn about a diagnostic test. Canadian Medical Association Journal, 124, 703-710.

Department of Clinical Epidemiology & Biostatistics McMaster University. (1981b). How To Read Clinical Journals: III. To learn the clinical course and prognosis of disease. Canadian Medical Association Journal, 124, 869-872.

Department of Clinical Epidemiology & Biostatistics McMaster University. (1981c). How To Read Clinical Journals: I. Why read them and how to start reading them critically. Canadian Medical Association Journal, 124, 555-558.

Department of Clinical Epidemiology & Biostatistics McMaster University. (1981d). How To Read Clinical Journals: V. To distinguish useful from useless or even harmful therapy. Canadian Medical Association Journal, 124, 1156-1162.

Department of Clinical Epidemiology & Biostatistics McMaster University. (1981e). How To Read Clinical Journals: IV. To determine aetiology or causation. Canadian Medical Association Journal, 124, 985-990.

Devisch, I., & Murray, S. J. (2009). "We hold these truths to be self-evident": deconstructing "evidence-based" medical practice. Journal of Evaluation in Clinical Practice, 15(6), 950-954.

Dimond, E. G., Kittle, C. F., & Crockett, J. E. (1960). Comparison of internal mammary artery ligation and sham operation for angina pectoris. American Journal of Cardiology, 5, 483-486.

Djulbegovic, B., Guyatt, G. H., & Ashcroft, R. E. (2009). Epistemologic Inquiries in Evidence-Based Medicine. Cancer Control, 16(2), 158-168.

Doll, R., & Peto, R. (1980). Randomised controlled trials and retrospective controls. British Medical Journal, 280, 44.

Druilhe, P., Brandicourt, O., Chongsuphajaisiddhi, T., & Berthe, J. (1988). Activity of a combination of three cinchona bark alkaloids against Plasmodium falciparum in vitro. Antimicrobial agents and chemotherapy, 32(2), 250-254.

Drummond, M. F., Richardson, W. S., O'Brien, B. J., Levine, M., & Heyland, D. (1997). Users' guides to the medical literature. XIII. How to use an article on economic analysis of clinical

practice. A. Are the results of the study valid? Evidence-Based Medicine Working Group. JAMA, 277(19), 1552-1557.

Echt, D. S., Liebson, P. R., Mitchell, L. B., Peters, R. W., Obias-Manno, D., Barker, A. H., Arensberg, D., et al. (1991). Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The Cardiac Arrhythmia Suppression Trial. New England Journal of Medicine, 324(12), 781-788.

Editorial. (2005). The end of homoeopathy. Lancet, 366, 690.

Edwards, A. G. K., Russell, I. T., & Stott, N. C. H. (1998). Signal versus noise in the evidence base for medicine: an alternative to hierarchies of evidence? Family Practice, 15(4), 319-322.

Eells, E., & Fitelson, B. (2001). Symmetries and Asymmetries in Evidential Support. Philosophical Studies, 107(2), 129-142.

Eichhorn, E. J., & Bristow, M. R. (2001). The Carvedilol Prospective Randomized Cumulative Survival (COPERNICUS) trial. Current Controlled Trials in Cardiovascular Medicine, 2(1), 20-23.

Enck, P., Klosterhalfen, S., Weimer, K., Horing, B., & Zipfel, S. (2011). The placebo response in clinical trials: more questions than answers. Philosophical Transactions of the Royal Society B, 366(1572), 1889-1895.

Ernst, E, Pittler, M. H., & Wider, B. (Eds.). (2006). The desktop guide to complementary and alternative medicine: an evidence-based approach. Edinburgh: Mosby.

Ernst, Edzard. (1998). Efficacy of Homeopathic Arnica: A systematic review of placebo-controlled clinical trials. Archives of Surgery, 133, 1187-1190.

Ernst, Edzard. (1999). Homeopathic prophylaxis of headaches and migraine? A systematic review. Journal of Pain and Symptom Management, 18(5), 353-7.

Ernst, Edzard. (2002). A systematic review of systematic reviews of homeopathy. British Journal of Clinical Pharmacology, 54(6), 577-582.

Ernst, Edzard. (2007). The truth about homeopathy. British Journal of Clinical Pharmacology, 65(2),163-4.

Ernst, Edzard. (2008). Complementary and alternative medicine: what the NHS should be funding? British Journal of General Practice, 58(548), 208-209.

Ernst, Edzard. (2009a). Homoeopathy and I. International Journal of Clinical Practice, 63(11), 1558-61.

Ernst, Edzard. (2009b). Jigsaw of Evidence. Complementary Therapies in Medicine, 17, 247.

Ernst, Edzard. (2010a). No to homeopathy placebo. Guardian. Retrieved February 22, 2010, from http://www.guardian.co.uk/commentisfree/2010/feb/22/science-homeopathy-clinical-trials?INTCMP=SRCH

Ernst, Edzard. (2010b). The ethics of British professional homoeopaths. International Journal of Clinical Practice, 64(2), 147-148.

Ernst, Edzard. (2011a). Homeopathy, non-specific effects and good medicine. Rheumatology, 50(6), 1007-1008.

Ernst, Edzard. (2011b). Pharmacists and homeopathic remedies. American Journal of Health-System Pharmacy, 68(6), 478.

Ernst, Edzard, & Resch, K.-L. (1995). Concept of true and perceived placebo effects. British Medical Journal, 311, 551-553.

Ernst, Edzard, Cohen, M. H., & Stone, J. (2004). Ethical problems arising in evidence based complementary and alternative medicine. Journal of Medical Ethics, 30(2), 156-159.

European Parliament. (2001). Directive 2001/83/EC. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2001:311:0067:0128:EN:PDF> Accessed 1/11/10.

European Parliament (1992). Directive 92/37/EC. <http://eur-lex.europa.eu/smartapi/cgi/sga_doc?smartapi!celexplus!prod!CELEXnumdoc&numdoc=392L0073&lg=en> Accessed 1/11/10.

Evans, I., Thornton, H., Chalmers, I., & Glasziou, P. (2011). Testing Treatments (Second.). London: Pinter & Martin.

Evans, S. J. (2009). Back to Square One. British Medical Journal, 338, 1404.

Everitt, B. S., & Rabe-Hesketh, S. (1997). The Analysis of Proximity Data. London: Arnold.

Evers, A. S., & Crowder, C. M. (2009). Mechanisms of Anesthesia and Consciousness. In P. G. Barash, B. F. Cullen, R. K. Stoelting, & M. Cahalan (Eds.), Clinical Anesthesia. London: Lippincott Williams & Wilkins.

Evidence Based Medicine Working Group. (1992). Evidence-Based Medicine: A new approach to the teaching the practice of medicine. JAMA, 268(17), 2420-2425.

Feder, G., & Katz, T. (2002). Randomised controlled trials for homoeopathy. British Medical Journal, 324(7336), 498-499.

Feinstein, A. R. (1964a). Scientific Methodology in Clinical Medicine II: Classification of human disease by clinical behaviour . Annals of Internal Medicine, 61(4), 757-781.

Feinstein, A. R. (1964b). Scientific Methodology in Clinical Medicine IV: Acquisition of clinical data . Annals of Internal Medicine, 61(6), 1162-1193.

Feinstein, A. R. (1964c). Scientific Methodology in Clinical Medicine I: Introduction, principles and concepts. Annals of Internal Medicine, 61(3), 564-579.

Feinstein, A. R. (1964d). Scientific Methodology in Clinical Medicine III: The evaluation of therapeutic response . Annals of Internal Medicine, 61(5), 944-965.

Feinstein, A. R. (1968a). Clinical Epidemiology I: The populational experiments of nature and man in human illness. Annals of Internal Medicine, 69(4), 807-820.

Feinstein, A. R. (1968b). Clinical Epidemiology III: The clinical design of statistics in therapy. Annals of Internal Medicine, 69(6), 1287-1311.

Feinstein, A. R. (1968c). Clinical Epidemiology II: The identification rates of disease. Annals of Internal Medicine, 69(5), 1037-1061.

Finniss, D. G., & Benedetti, F. (2005). Mechanisms of the placebo response and their impact on clinical trials and clinical practice. Pain, 114(1-2), 3-6.

Fisher, P. (2006). Homeopathy and The Lancet. Evidence-based Complementary and Alternative Medicine: eCAM, 3(1), 145-147.

Fisher, P. (2007). The Memory of Water: a scientific heresy? Homeopathy, 96(3), 141-142.

Fisher, P. (2009). The singer, the song, or both? Homeopathy, 98(2), 75-76.

Fisher, P. (2010). Ockham's Razor or Procrustes' Axe? Why we should reject philosophical speculation that ignores fact. Journal of Evaluation in Clinical Practice, 16(2), 282-3.

Fisher, P., Berman, B., Davidson, J., Reilly, D., & Thompson, T. D. B. (2005). Are the clinical effects of homoeopathy placebo effects? Lancet, 366, 2082-2083.

Flaten, M. A. (1998). Information about drug effects modify arousal: An investigation of the placebo response. Nordic Journal of Psychiatry, 52(2), 147-151.

Flaten, M. A. (2009). Drug effects: Agonistic and antagonistic processes. Scandinavian Journal of Psychology, 50(6), 652-659.

Flaten, M. A., Simonsen, T., & Olsen, H. (1999). Drug-Related Information Generates Placebo and Nocebo Responses That Modify the Drug Response. Psychosomatic Medicine, 61, 250-255.

Flaten, M. A., Simonsen, T., Zahlsen, K., Aamo, T., Sager, G., & Olsen, H. (2004). Stimulant and relaxant drugs combined with stimulant and relaxant information: A study of active placebo. Psychopharmacology, 176(3), 426-434.

Fletcher, R. H., Fletcher, S. W., & Wagner, E. H. (1982). Clinical Epidemiology: The Essentials. Baltimore: Williams & Wilkins.

Fletcher, S., Wagner, E., & Fletcher, R. (1996). Clinical Epidemiology: The Essentials. London: Williams & Wilkins.

Foddy, B. (2009a). A Duty to Deceive: Placebos in Clinical Practice. American Journal of Bioethics, 9(12), 4-12.

Foddy, B. (2009b). Response to open peer commentaries on "A duty to deceive: placebos in clinical practice." American Journal of Bioethics, 9(12), W1-2.

Frass, M., Schuster, E., Muchitsch, I., Duncan, J., Gei, W., Kozel, G., Kastinger-Mayr, C., et al. (2005). Bias in the trial and reporting of trials of homeopathy: a fundamental breakdown in peer review and standards? Journal of Alternative & Complementary Medicine, 11(5), 780–782.

Freeman, H. (2010). Me and my homeopathic overdose. Guardian. Retrieved February 3, 2010, from http://www.guardian.co.uk/commentisfree/2010/feb/03/homeopathy-overdose-hadley-freeman

Furnham, A. (1996). Why do People Choose and Use Complementary Therapies? In E Ernst (Ed.), Complementary Medicine: An Objective Appraisal. Oxford: Butterworth Heinemann.

GH, G., Sackett, D., Sinclair, JC, Haywood, R., Cook, D., & Cook, R. J. (1995). Users' guides to the medical literature, IX: a method for grading health care recommendations. JAMA, 274, 1800-1804.

Garattini, S., & Bertelé, V. (2009). Homoeopathy: not a matter for drug-regulatory authorities. Lancet, 374(9701), 1578-1580.

Gerber, A., & Lauterbach, K. W. (2005). Evidence-based medicine: why do opponents and proponents use the same arguments? Health Care Analysis, 13(1), 59-71.

Giacomini, M. K., & Cook, D. J. (2000a). Users' Guides to the Medical Literature: XXIII. Qualitative Research in Health Care B. What Are the Results and How Do They Help Me Care for My Patients? JAMA, 284(4), 478-482.

Giacomini, M. K., & Cook, D. J. (2000b). Users' Guides to the Medical Literature: XXIII. Qualitative Research in Health Care A. Are the Results of the Study Valid? JAMA, 284(3), 357-362.

Glasziou, P., Chalmers, I., Rawlins, M., & McCulloch, P. (2007). When are randomised trials unnecessary? Picking signal from noise. British Medical Journal, 334(7589), 349-351.

Glennan, S. (1996). Mechanisms and the Nature of Causation. Erkenntnis, 44, 49-71.

Glennan, S. (2002). Rethinking Mechanistic Explanation. Philosophy of Science, 69, S342-353.

Gold, P. W., Novella, S, Roy, R. R., Marcus, D., Bell, I. R., Davidovitch, N, & Saine, A. (2008). Homeopathy—quackery or a key to the future of medicine? Homeopathy, 97(1), 28-33.

Goldacre, B. (2007a). A kind of magic? Guardian. Retrieved November 16, 2007, from http://www.guardian.co.uk/science/2007/nov/16/sciencenews.g2

Goldacre, B. (2007b). Benefits and risks of homoeopathy. Lancet, 370(9600), 1672-1673.

Goldacre, B. (2008a). Don't laugh, sugar pills are the future. Guardian. Retrieved March 1, 2008, from http://www.guardian.co.uk/science/2008/mar/01/medicalresearch.health

Goldacre, B. (2008b). Bad Science. London: Fourth Estate.

Goldacre, B. (2011). Bad Science Blog. <http://www.badscience.net/>

Goldenberg, M. J. (2006). On evidence and evidence-based medicine: Lessons from the philosophy of science. Social Science & Medicine, 62(11), 2621-2632.

Goldenberg, M. J., Borgerson, K., & Bluhm, R. (2009). The Nature of Evidence in Evidence-Based Medicine: Guest editors' Introduction. Perspectives in Biology and Medicine, 52(2), 164-167.

Goldhammer, D. H. (1969). Toward a more General Inquirer: convergence of structure. The analysis of communication content. New York: John Wiley & Sons.

Goldman, N., Chen, M., Fujita, T., Xu, Q., Peng, W., Liu, W., Jensen, T. K., et al. (2010). Adenosine A1 receptors mediate local anti-nociceptive effects of acupuncture. Nature Neuroscience, 13(7), 883-888.

Golomb, B. A. (1995). Paradox of placebo effect. Nature, 375, 530.

Golomb, B. A., Erickson, L., Koperski, S., Sack, D., Enkin, M. W., & Howick, J. (2010). What's in Placebos: Who Knows? Analysis of Randomized, Controlled Trials. Annals of Internal Medicine, 153, 532-536.

Google. (2010). Homeopathy 1800-2000 ngram. Retrieved December 24, 2010, from http://goo.gl/jIE3M

Grade Working Group. (2004). Grading quality of evidence and strength of recommendations. British Medical Journal, 328(7454), 1490-1494.

Greenhalgh, Trisha. (2002). Intuition and evidence - uneasy bedfellows? British Journal of General Practice, 52(478), 395-400.

Greenhalgh, Trisha. (2006). How to read a paper: the basics of evidence-based medicine (Third Edition). Oxford: BMJ Books.

Grevert, P., & Goldstein, A. (1978). Endorphins: naloxone fails to alter experimental pain or mood in humans. Science, 199(4333), 1093-1095.

Grossman, J., & Mackenzie, F. J. (2005). The randomized controlled trial: gold standard, or merely standard? Perspectives in Biology and Medicine, 48(4), 516-534.

Grünbaum, A. (1981). The Placebo Concept. Behavioural Research & Therapy, 19(2), 157-167.

Grünbaum, A, Cicchetti, D., & Grove, W. M. (1991). The Placebo Concept in Medicine and Psychiatry. Oxford: University of Minnesota Press.

Grünbaum, Adolf. (1986). The placebo concept in medicine and psychiatry. Psychological Medicine, 16(1), 19-38.

Guajardo, G., & Wilson, J. (2005). Models for explaining the homeopathic healing process: a historical and critical account of principles central to homeopathy. Homeopathy, 94, 44-48.

Gupta, M. (2003). A critical appraisal of evidence-based medicine: some ethical considerations. Journal of Evaluation in Clinical Practice, 9(2), 111-121.

Guyatt, G. H. (1991). Evidence Based Medicine. Annals of Internal Medicine, suppl. 2: ACP Journal Club, 114, A-16.

Guyatt, G. H., & Drummond, M. F. (2002a). Users' Guide to the Medical Literature: Essentials for Evidence-Based Clinical Practice. Chicago: AMA Press.

Guyatt, G. H., & Drummond, M. F. (2002b). Users' Guide to the Medical Literature: A Manual for Evidence-Based Clinical Practice. Chicago: AMA Press.

Guyatt, G. H., & Rennie, D. (1993). Users' guides to the medical literature. JAMA, 270(17), 2096-2097.

Guyatt, G. H., & Rennie, Drummond (Eds.). (2002). Users' guides to the medical literature: essentials of evidence-based clinical practice (First Edition). Chicago: AMA Press.

Guyatt, G. H., Haynes, R Brian, Jaeschke, R. Z., Cook, D. J., Green, L., Naylor, C. D., Wilson, M. C., et al. (2000). Users' Guides to the Medical Literature: XXV. Evidence-Based Medicine: Principles for Applying the Users' Guides to Patient Care. JAMA, 284(10), 1290-1296.

Guyatt, G. H., Naylor, C. D., Juniper, E., Heyland, D. K., Jaeschke, R. Z., & Cook, D. J. (1997). Users' guides to the medical literature. XII. How to use articles about health-related quality of life. Evidence-Based Medicine Working Group. JAMA, 277(15), 1232-1237.

Guyatt, G. H., Oxman, A. D., Kunz, R., Falck-Ytter, Y., Vist, G. E., Liberati, A., & Schünemann, H. J. (2008). GRADE: going from evidence to recommendations. British Medical Journal, 336(7658), 1049-1051 [Erratum BMJ 336 doi: 10.1136/bmj.a402].

Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., & Schünemann, H. J. (2008). GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. British Medical Journal, 336, 924-926.

Guyatt, G. H., Sinclair, Jack, Cook, D. J., & Glasziou, P. (1999). Users' Guides to the Medical Literature: XVI. How to Use a Treatment Recommendation. JAMA, 281(19), 1836-1843.

Guyatt, G., Sackett, D., & Cook, D. (1993). Users' guides to the medical literature, II: how to use an article about therapy or prevention, A: are the results of the study valid? JAMA, 270, 2598-2601.

Guyatt, G., Sackett, D., & Cook, D. (1994). Users' guides to the medical literature, II: how to use an article about therapy or prevention, B: what were the results and will they help me in caring for my patients? JAMA, 271, 59-63.

Gøtzsche, P. C. (1994). Is there logic in the placebo? Lancet, 344(8927), 925-926.

Gøtzsche, P. C. (1995). Placebo effects. Concept of placebo should be discarded. British Medical Journal, 311(7020), 1640.

Haake, M., Müller, H.-H., Schade-Brittinger, C., Basler, H. D., Schäfer, H., Maier, C., Endres, H. G., et al. (2007). German Acupuncture Trials (GERAC) for chronic low back pain: randomized, multicenter, blinded, parallel-group trial with 3 groups. Archives of Internal Medicine, 167(17), 1892-1898.

Hahn, R. A., & Harrington, A. (1997). The Nocebo Phenomenon: Scope and Foundations. London: Harvard University Press.

Hahn, R. A., & Kleinman, A. (1983). Belief as Pathogen, Belief as Medicine: "Voodoo Death" and the "placebo Phenomenon" in Anthropological Perspective. Medical Anthropology Quarterly, 14(4), 3-19.

Hahnemann, S. (1805). Fragmenta de viribus medicamentorum positivissive in sano corpore observatis. Leipzig: Barth.

Hahnemann, S. (1852). The Lesser Writings of Samuel Hahnemann. (R. E. Dudgeon, Ed.). New York: Raddle.

Hahnemann, S. (1904). The Chronic Diseases. (L. H. Tafel, Ed.). Philadelipha: Boericke.

Hahnemann, S. (1983). Organon of Medicine. (J. Künzli, A. Naudé, & P. Pendleton, Eds.) (6th ed.). London: Gollancz.

Hansen, K., & Kappel, K. (2010). The proper role of evidence in complementary/alternative medicine. Journal of Medicine and Philosophy, 35(1), 7-18.

Harbour, R., Miller, J., & SIGN Grading Review Group. (2001). A new system for grading recommendations in evidence based guidelines. British Medical Journal, 323, 334-336.

Harrington, A. (1997). Introduction. In A Harrington (Ed.), The Placebo Effect: An Interdisciplinary Exploration. London: Harvard University Press.

Harrington, Anne. (2002). “seeing” the placebo effect: historical legacies and present opportunities. In H. A. Guess, A. Kleinman, J. W. Kusek, & L. W. Engel (Eds.), The Science of Placebo: Toward an interdisciplinary research agenda. London: BMJ Books.

Harris, E. (2011). Is homeopathy on the ropes after ban on prescription for pets? Guardian. Retrieved January 5, 2011, from http://www.guardian.co.uk/science/political-science/2011/jan/05/homeopathy-ban- prescription-pets

Hay, L. (2008). Will new regulations reverse the “drop ” in homeopathy? Perspectives in Public Health, 128(6), 282-283.

Haynes, R Brian. (2002). What kind of evidence is it that Evidence-Based Medicine advocates want health care providers and consumers to pay attention to? BMC Health Services Research, 2(1), 3.

Hays, D. (1969). Linguistic foundations for a theory of content analysis. The analysis of communication content. New York: John Wiley & Sons.

Hayward, R. S., Wilson, M. C., Tunis, S. R., Bass, E. B., & Guyatt, G. H. (1995). Users’ guides to the medical literature. VIII. How to use clinical practice guidelines. A. Are the recommendations valid? The Evidence-Based Medicine Working Group. JAMA, 274(7), 570-574.

Hempel, C. G. (1968). Maximal Specificity and Lawlikeness in Probabilistic Explanation. Philosophy of Science, 35(2), 116-133.

Hester, D. M., & Talisse, R. B. (2009). Physician deception and patient autonomy. American Journal of Bioethics, 9(12), 22-23.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. British Medical Journal, 327, 557-560.

Hill, A. B. (1951). The Clinical Trial. British Medical Bulletin, 7(4), 278-282.

Hill, A. B. (1965). The Environment and Disease: Association or Causation? Proceedings of the Royal Society of Medicine, 58(5), 295-300.

Hoffer, L. J. (2003). Complementary or alternative medicine: the need for plausibility. Canadian Medical Association Journal, 168(2), 180-182.

Holmes, O. W. (1842). Homeopathy and its Kindred Delusions. Retrieved December 2010, from http://ebooks.adelaide.edu.au/h/holmes/oliver_wendell/homeopathy/index.html

Holt, S., Gilbey, A., Colquhoun, D., Baum, M., & Ernst, Edzard. (2011). Call for doctors not to practice homeopathy or refer to homeopaths. New Zealand Medical Journal, 124(1332), 3-4.

House of Commons Science & Technology Committee. (2010). Evidence Check 2: Homeopathy. London. <http://www.publications.parliament.uk/pa/cm200910/cmselect/cmsctech/45/4502.htm>

Howick, J. (2009). Questioning the Methodologic Superiority of "Placebo" Over "Active" Controlled Trials. American Journal of Bioethics, 9(9), 34-48.

Howick, J. (2011). The Philosophy of Evidence-Based Medicine. Oxford: Wiley-Blackwell.

Howick, J., Chalmers, I., Glasziou, P., Greenhalgh, Trish, Heneghan, C., Liberati, A., Moschetti, I., Phillips, B., & Thornton, H. (2011a). Explanation of the 2011 Oxford Centre for Evidence-Based Medicine (OCEBM) Levels of Evidence: Background document. Oxford Centre for Evidence-Based Medicine. Retrieved January 20, 2011, from http://www.cebm.net/index.aspx?o=5653

Howick, J., Chalmers, I., Glasziou, P., Greenhalgh, Trish, Heneghan, C., Liberati, A., Moschetti, I., Phillips, B., & Thornton, H. (2011b). The 2011 Oxford CEBM Levels of Evidence: Introductory Document. Oxford Centre for Evidence-Based Medicine. Retrieved January 20, 2011, from http://www.cebm.net/index.aspx?o=5653

Howick, J., Glasziou, P., & Aronson, J. K. (2009). The evolution of evidence hierarchies: what can Bradford Hill's "guidelines for causation" contribute? Journal of the Royal Society of Medicine, 102(5), 186-194.

Hróbjartsson, A., & Gøtzsche, P. C. (2001). Is The Placebo Powerless? An Analysis of Clinical Trials Comparing Placebo with No Treatment. New England Journal of Medicine, 344(21), 1594-1603.

Hróbjartsson, A., & Gøtzsche, P. C. (2004). Is the placebo powerless? Update of a systematic review with 52 new randomized trials comparing placebo with no treatment. Journal of Internal Medicine, 256, 91-100.

Hróbjartsson, A., & Gøtzsche, P. C. (2010). Placebo interventions for all clinical conditions. Cochrane Database of Systematic Reviews, (1), CD003974. doi:10.1002/14651858.CD003974.pub3

Hróbjartsson, A., Kaptchuk, T. J., & Miller, F. G. (2011). Placebo effect studies are susceptible to response bias and to other types of biases. Journal of Clinical Epidemiology, 64(11), 1223-1229.

Hughes, J. R., Gulliver, S. B., Amori, G., Mireanlt, G. C., & Fenwick, J. F. (1989). Effect of instructions and nicotine on smoking cessation, withdrawal symptoms and self-administration of nicotine gum. Psychopharmacology, 99(4), 486-491.

Hunt, D. L., Jaeschke, R. Z., & McKibbon, K. A. (2000). Users' Guides to the Medical Literature: XXI. Using Electronic Health Information Resources in Evidence-Based Practice. JAMA, 283(14), 1875-1879.

Hunter, M. (2002). Homeopathy use in NHS not justified. British Medical Journal, 324, 565.

Illich, I. (1976). Limits to Medicine. Medical nemesis: the expropriation of health. London: Boyars.

Imam, I. (2011). Cognitive biases affect clinical judgment. British Medical Journal, 343, d7705.

Innovation Universities Science and Skills Committee Press Release. (2009). MPs call for Government to establish a House of Commons Science and Technology Committee following DIUS/BERR merger. <http://goo.gl/F6Wi >

Ioannidis, J. P. A. (2005). Why most published research findings are false. PLoS medicine, 2(8), e124. doi:10.1371/journal.pmed.0020124

Jaeschke, R. Z., Guyatt, G. H., & Sackett, D. L. (1994a). Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. JAMA, 271(9), 703-707.

Jaeschke, R. Z., Guyatt, G. H., & Sackett, D. L. (1994b). Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? Evidence-Based Medicine Working Group. JAMA, 271(5), 389-391.

Jameslindlibrary.org. (2011). Forbes J (1846). Homoeopathy, allopathy and "young physic". British and Foreign Medical Review 225-265. James Lind Library. Retrieved July 9, 2011, from http://www.jameslindlibrary.org/illustrating/records/homoeopathy-allopathy-and-young-physic/images.pdf

Jenicek, M. (2006). Evidence-based medicine: Fifteen years later. Golem the good, the bad and the ugly in need of a review? Medical Science Monitor, 12(11), 241-252.

Johnson, A. G. (1994). Surgery as a placebo. Lancet, 140, 142.

Jonas, W. B. (2001). The Evidence House. Western Journal of Medicine, 175(2), 79-80.

Jonas, W. B., Kaptchuk, T. J., & Linde, K. (2003). A critical overview of homeopathy. Annals of Internal Medicine, 138(5), 399.

Jones, M. (2006). Malaria advice "risks lives." Newsnight. Retrieved August 2008, from http://news.bbc.co.uk/1/hi/programmes/newsnight/5178122.stm

Jones, M., & Ghosh, P. (2011). Cases dropped against malaria homeopaths. BBC News. Retrieved January 12, 2011, from http://www.bbc.co.uk/news/health-12153074

Kaptchuk, T. J. (2001). The double-blind, randomized, placebo-controlled trial: Gold standard or golden calf? Journal of Clinical Epidemiology, 54, 541-549.

Kaptchuk, T. J. (2002). Academia and Clinic The Placebo Effect in Alternative Medicine : Can the Performance of a Healing Ritual Have Clinical Significance ? Annals of Internal Medicine, (7), 817-825.

Kaptchuk, T. J. (2011). Placebo studies and ritual theory: a comparative analysis of Navajo, acupuncture and biomedical healing. Philosophical Transactions of the Royal Society B, 366(1572), 1849-1858.

Kaptchuk, T. J., Friedlander, E., Kelley, J. M., Sanchez, M. N., Kokkotou, E., Singer, J. P., Kowalczykowski, M., et al. (2010). Placebos without deception: a randomized controlled trial in irritable bowel syndrome. PLoS ONE, 5(12), e15591. doi:10.1371/journal.pone.0015591

Kaptchuk, T. J., Goldman, P., Stone, D. A., & Stason, W B. (2000). Do medical devices have enhanced placebo effects. Journal of Clinical Epidemiology, 53(8), 786-792.

Kaptchuk, T. J., Stason, William B, Davis, R. B., Legedza, A. R. T., Schnyer, R. N., Kerr, C. E., Stone, D. a, et al. (2006). Sham device v inert pill: randomised controlled trial of two placebo treatments. British Medical Journal, 332, 391-397.

Katz, D. (2001). Clinical Epidemiology & Evidence-Based Medicine. London: Sage.

Kaufman, M. (1971). Homeopathy in America: The rise and fall of a medical heresy. Baltimore: The Johns Hopkins Press.

Kavalier, F. (2011). Why do pharmacists sell homoeopathic remedies? British Medical Journal, 342, d349.

Keshet, Y. (2009). The untenable boundaries of biomedical knowledge: epistemologies and rhetoric strategies in the debate over evaluating complementary and alternative medicine. Health, 13(2), 131-155.

Kiene, H., Kienle, G. S., & von Schön-Angerer, T. (2005). Failure to exclude false negative bias: a fundamental flaw in the trial of Shang et al. Journal of Alternative & Complementary Medicine, 11(5), 783.

Kienle, G. S., & Kiene, H. (1997). The powerful placebo effect: Fact or fiction. Journal of Clinical Epidemiology, 50(12), 1311-1318.

Kirmayer, L. J. (1993). Healing and the invention of metaphor: the effectiveness of symbols revisited. Culture, Medicine and Psychiatry, 17(2), 161-195.

Kirmayer, L. J. (2004). The cultural diversity of healing: meaning, metaphor and mechanism. British Medical Bulletin, 69(1), 33-48.

Kirsch, I. (1985). Response expectancy as a determinant of experience and behavior. American Psychologist, 40(11), 1189-1202.

Kirsch, I. (2000). Are drug and placebo effects in depression additive? Biological Psychiatry, 47(8), 733-735.

Kirsch, I. (2011). Preface. Philosophical Transactions of the Royal Society B, 366(1572), 1781-1782.

Kirsch, I., Deacon, B. J., Huedo-medina, T. B., Scoboria, A., Moore, T. J., & Johnson, B. T. (2008). Initial Severity and Antidepressant Benefits: A Meta-analysis of Data Submitted to the Food and Drug Administration. PLoS Med, 5(2), e45.

Kleijnen, J., Craen, A. J. D., Everdingen, J. V., & Krol, L. (1994). Placebo Effect in Double-Blind Clinical Trials: A Review of Interactions with Medications. Lancet, 344, 1347-1349.

Kleijnen, J., Knipschild, P., & ter Riet, G. (1991). Clinical trials of homoeopathy. British Medical Journal, 302, 316-323.

Kleinman, A. (1977). Lessons from a Clinical Approach to Medical Anthropological Research. Medical Anthropology Newsletter, 8(4), 11-15.

Kleinman, A., Mendelsohn, E., & Elkana, Y. (1981). The Meaning Context of Illness and Care: Reflections on a central theme in the anthropolgy of medicine. Dordrecht: Reidel.

Koshi, E. B., & Short, C. A. (2007). Placebo theory and its implications for research and clinical practice: A review of the recent literature. Pain Practice, 7, 4-20.

Kristiansen, I. S., & Mooney, G. (2004). Evidence-Based Medicine: In its place. London: Routledge.

Kruskal, J. B., & Wish, M. (1978). Multidimensional Scaling. London: Sage Publications.

Lambert, H. (2006). Accounting for EBM: notions of evidence in medicine. Social Science & Medicine, 62(11), 2633-2645.

Laupacis, A, Wells, G. A., Richardson, W. S., & Tugwell, P. (1994). Users' guides to the medical literature. V. How to use an article about prognosis. Evidence-Based Medicine Working Group. JAMA, 272(3), 234-237.

Leckridge, B. (1997). Homeopathy in Primary Care. London: Churchill-Livingstone.

Leckridge, B. (2008). The diverse nature of homeopathic practice. In S. B. Kayne (Ed.), Homeopathic Practice. London: Pharmaceutical press.

Leibovici, L. (2001). Effects of remote, retroactive intercessory prayer on outcomes in patients with bloodstream infection: randomised controlled trial. British Medical Journal, 323(7327), 1450-1451.

Levine, J D, Gordon, N C, Smith, R, & Fields, H L. (1981). Analgesic responses to morphine and placebo in individuals with postoperative pain. Pain, 10(3), 379-389.

Levine, Jon D, Gordon, Newton C, & Fields, Howard L. (1978). The Mechanism of Placebo Analgesia. Lancet, 2, 654-657.

Levine, M., Walter, S., Lee, H., Haines, T., Holbrook, A., & Moyer, V. A. (1994). Users' guides to the medical literature. IV. How to use an article about harm. Evidence-Based Medicine Working Group. JAMA, 271(20), 1615-1619.

Lewis, A. (2010). An Obituary: Royal London Homeopathic Hospital, 1849-2010. The Quackometer. Retrieved October 2010, from http://www.quackometer.net/blog/2010/09/an-obituary-royal-london-homeopathic-hospital-1849-2010.html

Lewis, A. (2011). The Quackometer blog. <http://quackometer.net/>

Lichtenberg, P., Heresco-Levy, U., & Nitzan, U. (2004). The ethics of the placebo in clinical practice. Journal of Medical Ethics, 30(6), 551-554.

Linde, K., & Jonas, W. B. (2005). Are the clinical effects of homoeopathy placebo effects? Lancet, 366(9503), 2081-2082.

Linde, K., & Melchart, D. (1998). Randomized Controlled Trials of Individualized Homeopathy: A State-of-the-Art Review. Journal of Alternative & Complementary Medicine, 4(4), 371-388.

Linde, K., Clausius, N., Ramirez, G., Melchart, D., Eitel, F., Hedges, L. V., & Jonas, W. B. (1997). Are the clinical effects of homoeopathy placebo effects? A meta-analysis of placebo-controlled trials. Lancet, 350, 834-843.

Linde, K., Fässler, M., & Meissner, K. (2011). Placebo interventions, placebo effects and clinical practice. Philosophical Transactions of the Royal Society B, 366(1572), 1905-1912.

Linde, K., Scholz, M., Ramirez, G., Clausius, N., Melchart, D., & Jonas, W. B. (1999). Impact of study quality on outcome in placebo-controlled trials of homeopathy. Journal of Clinical Epidemiology, 52(7), 631-636.

Lipsett, A. (2008). Homeopathy degree suspended after criticism. Guardian. Retrieved August 27, 2007, from http://www.guardian.co.uk/education/2008/aug/27/highereducation.research

Lipsett, A. (2009). The opposite of science. Guardian. Retrieved February 24, 2009, from http://www.guardian.co.uk/education/2009/feb/24/homeopathy-science

Long, L., & Ernst, Edzard. (2001). Homeopathic remedies for the treatment of osteoarthritis: a systematic review. British Homoeopathic Journal, 90(1), 37-43.

Loughlin, M. (2009). The basis of medical knowledge: judgement, objectivity and the history of ideas. Journal of Evaluation in Clinical Practice, 15(6), 935-940.

Louhiala, P., & Puustinen, R. (2008). Rethinking the placebo effect. British Medical Journal, 34(2), 107-109.

Lüdtke, R., & Rutten, A. L. B. (2008). The conclusions on the effectiveness of homeopathy highly depend on the set of analyzed trials. Journal of Clinical Epidemiology, 61(12), 1197-1204.

MHRA Consultation Letter MLX 312. (2005). <http://goo.gl/t753Z>: accessed 1/11/10.

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking About Mechanisms. Philosophy of Science, 67(1), 1-25.

Martin, E. A. (Ed.). (2007). Oxford Concise Colour Medical Dictionary (Fourth Edition). Oxford: Oxford University Press.

Mashta, O. (2009). WHO warns against using homoeopathy to treat serious diseases. British Medical Journal, 339, b3447.

McAlister, F. A., Laupacis, Andreas, Wells, G. A., Sackett, D. L., & for the Evidence-Based Medicine Working, G. (1999). Users' Guides to the Medical Literature: XIX. Applying Clinical Trial Results; B. Guidelines for Determining Whether a Drug Is Exerting (More Than) a Class Effect. JAMA, 282(14), 1371-1377.

McAlister, F. A., Straus, S. E., Guyatt, G. H., & Haynes, R Brian. (2000). Users' Guides to the Medical Literature: XX. Integrating Research Evidence With the Care of the Individual Patient. JAMA, 283(21), 2829-2836.

McCarney, R. W., Lasserson, T. J., Linde, K., & Brinkhaus, B. (2004). An overview of two Cochrane systematic reviews of complementary treatments for chronic asthma: acupuncture and homeopathy. Respiratory Medicine, 98(8), 687-696.

McCarney, R. W., Warner, J., Fisher, P., & van Haselen, R. (2003). Homeopathy for dementia (Review). Cochrane Database of Systematic Reviews, (1), 1-10. doi:10.1002/14651858.CD003803

McGinn, T. G., Guyatt, G. H., Wyer, P. C., Naylor, C. D., Stiell, I. G., & Richardson, W. S. (2000). Users' Guides to the Medical Literature: XXII: How to Use Articles About Clinical Decision Rules. JAMA, 284(1), 79-84.

Medical Research Council. (2000). A Framework for Development and Evaluation of RCTs for Complex Interventions to Improve Health.

Medicines Act. (1968) <http://www.legislation.gov.uk/ukpga/1968/67/contents> accessed 1/11/10.

Meissner, K., Kohls, N., & Colloca, Luana. (2011). Introduction to placebo effects in medicine: mechanisms and clinical implications. Philosophical Transactions of the Royal Society B, 366(1572), 1783-1789.

Meunier, P. J., Sebert, J., Reginster, J., Briancon, D., Appelboom, T., Netter, P., & Loeb, G. (1998). International Original Article Fluoride Salts are no Better at Preventing New Vertebral Fractures than Calcium – Vitamin D in Postmenopausal Osteoporosis: The FAVO Study. Osteoporosis International, 8, 4-12.

Middleton, H., & Moncrieff, Joanna. (2011). They won't do any harm and might do some good: time to think again on the use of antidepressants? British Journal of General Practice, 61(582), 47–49.

Milazzo, S., Russell, N., & Ernst, Edzard. (2006). Efficacy of homeopathic therapy in cancer treatment. European Journal of Cancer, 42(3), 282-289.

Milgrom, L. R. (2005). Are randomized controlled trials (RCTs) redundant for testing the efficacy of homeopathy? A critique of RCT methodology based on entanglement theory. Journal of Alternative & Complementary Medicine, 11(5), 831–838.

Milgrom, L. R. (2006a). Is homeopathy possible? Journal of the Royal Society for the Promotion of Health, 126(5), 211-218.

Milgrom, L. R. (2006b). Entanglement, knowledge, and their possible effects on the outcomes of blinded trials of homeopathic provings. Journal of Alternative & Complementary Medicine, 12(3), 271–279.

Milgrom, L. R. (2007). Conspicuous by its absence: the Memory of Water, macro-entanglement, and the possibility of homeopathy. Homeopathy, 96(3), 209-219.

Milgrom, L. R. (2008). Homeopathy and the new fundamentalism: a critique of the critics. Journal of Alternative & Complementary Medicine, 14(5), 589–594.

Milgrom, L. R. (2009a). Under Pressure: Homeopathy UK and Its Detractors. Forschende Komplementärmedizin, 16(4), 256-261.

Milgrom, L. R. (2009b). Gold Standards, Golden Calves, and Random Reproducibility: Why Homeopaths at Last Have Something to Smile About. Journal of Alternative & Complementary Medicine, 15(3), 205-207.

Milgrom, L. R. (2009c). Falling Trees, Fractal , and Sophistry: Some Philosophical “Biohazards” En Route to Reconciling Biomedicine and Homeopathy. Journal of Alternative & Complementary Medicine, 15(11), 1247-1254.

Milgrom, L. R. (2010a). Response to Leslie Rose and John Garrow: Clinical trials, clinical evidence, and selective citation Misunderstanding RCTs is all “tu quoque.” Homeopathy, 99(2), 149-150.

Milgrom, L. R. (2010b). When sorry seems to be the hardest word: CAM, free speech, and the British legal system. Homeopathy, 99(1), 83-84.

Miller, F. G., & Brody, H. (2011). Understanding and Harnessing Placebo Effects: Clearing Away the Underbrush [EARLY VIEW]. Journal of Medicine and Philosophy, 1-10. doi:10.1093/jmp/jhq061

Miller, F. G., & Colloca, Luana. (2009). The legitimacy of placebo treatments in clinical practice: evidence and ethics. American Journal of Bioethics, 9(12), 39-47.

Miller, F. G., & Colloca, Luana. (2011). The placebo phenomenon and medical ethics: rethinking the relationship between informed consent and risk-benefit assessment. Theoretical Medicine and Bioethics, 32(4), 229-243.

Miller, F. G., & Kaptchuk, T. J. (2008). The power of context: reconceptualizing the placebo effect. Journal of the Royal Society of Medicine, 101(5), 222-5.

Mitchell, S. H., Laurent, C. L., & Wit, H. (1996). Interaction of expectancy and the pharmacological effects of d-amphetamine: subjective effects and self-administration. Psychopharmacology, 125(4), 371-378.

Moerman, D. E. (1983). General Medical Effectiveness and Human Biology: Placebo Effects in the Treatment of Ulcer Disease. Medical Anthropology Quarterly, 14(4), 3+13-16.

Moerman, D. E. (2000). Cultural Variations in the Placebo Effect: Ulcers, Anxiety, and Blood Pressure. Medical Anthropology Quarterly, 14(1), 51-72.

Moerman, D. E. (2002a). The Meaning Response and the Ethics of Avoiding Placebos. Evaluation & The Health Professions, 25(4), 399-409.

Moerman, D. E. (2002b). Meaning, Medicine, and the "Placebo Effect." Cambridge: Cambridge University Press.

Moerman, D. E., & Jonas, W. B. (2002). Deconstructing the Placebo Effect and Finding the Meaning Response. Annals of Internal Medicine, 136(6), 471-476.

Moerman, D. E., Benoist, J., Brody, E. B., Giovannini, M., Gracia, M. F., Hall, E. T., Heggenhougen, H. K., et al. (1979). Anthropology of Symbolic Healing [and Comments and Reply]. Current Anthropology, 20(1), 59-80.

Moncrieff, J, Wessely, S., & Hardy, R. (2004). Active placebos versus antidepressants for depression. Cochrane Database of Systematic Reviews, (1), CD003012. doi:10.1002/14651858.CD003012.pub2

Moncrieff, Joanna. (2009). The Myth of the Chemical Cure: A critique of psychiatric drug treatment (Revised Ed.). Basingstoke: Pelgrave Macmillan.

Montori, V. M., & Guyatt, G. H. (2008). Progress in evidence-based medicine. JAMA, 300(15), 1814-1816.

Mora, M. S., Nestoriuc, Y., & Rief, W. (2011). Lessons learned from placebo groups in antidepressant trials. Philosophical Transactions of the Royal Society B, 366(1572), 1879-1888.

Moynihan, R., & Smith, Richard. (2002). Too much medicine? British Medical Journal, 324, 859-860.

Mumford, S., & Anjum, R. L. (2011). Getting Causes from Powers. Oxford University Press: Oxford.

NHS. (2010). Homeopathy remains on NHS. NHS Choices: News. Retrieved July 27, 2010, from http://www.nhs.uk/news/2010/July07/Pages/nhs-homeopathy.aspx

Naylor, C. D., & Guyatt, G. H. (1996a). Users' guides to the medical literature. X. How to use an article reporting variations in the outcomes of health services. The Evidence-Based Medicine Working Group. JAMA, 275(7), 554-558.

Naylor, C. D., & Guyatt, G. H. (1996b). Users' guides to the medical literature. XI. How to use an article about a clinical utilization review. Evidence-Based Medicine Working Group. JAMA, 275(18), 1435-1439.

Newell, S. J. (2011). Doctors must bear in mind the fallibility of their judgement. British Medical Journal, 343, d7709.

Nicholls, P. (1988). Homoeopathy and the Medical Profession. London: Croom Helm.

Nicolai, T. (2008). Important concepts and the approach to prescribing. In S. B. Kayne (Ed.), Homeopathic Practice. London: Pharmaceutical Press.

Norman, G. R. (2001). Examining the assumptions of evidence-based medicine. Journal of Evaluation in Clinical Practice, 5(2), 139-147.

Novella, Steven. (2011). Science-Based Medicine blog. <http://www.sciencebasedmedicine.org>

Nunn, R. (2009a). It's time to put the placebo out of our misery. British Medical Journal, 338, b1568.

Nunn, R. (2009b). Preparing for a Post-Placebo Paradigm: Ethics and Choice of Control in Clinical Trials. American Journal of Bioethics, 9(9), 51-52.

Nunn, R. (2009c). Placebo effects without placebos? More reason to abandon the paradoxical placebo. American Journal of Bioethics, 9(12), 50-52.

OCEBM Levels of Evidence Working Group. (2009). The Oxford Levels of Evidence. Retrieved March 3, 2010, from http://www.cebm.net/index.aspx?o=1025

OCEBM Levels of Evidence Working Group. (2011). The Oxford Levels of Evidence. Retrieved January 20, 2011, from http://www.cebm.net/index.aspx?o=5653

Oberbaum, M., Singer, S. R., & Frass, M. (2005). Homeopathic research after the Lancet meta analysis — A moment for introspection. Complementary Therapies in Medicine, 13, 303-305.

Oberbaum, M., Vithoulkas, G., van Haselen, R., & Singer, S. R. (2003). Reinventing the wheel? Or the Emperor's new clothes. Journal of Alternative & Complementary Medicine, 9(5), 613-616.

Ong, L. M. L., De Haes, J., Hoos, A. M., & Lammes, F. B. (1995). Doctor-patient communication: a review of the literature. Social Science & Medicine, 40(7), 903-918.

Overall, K. L., & Dunham, A. E. (2009). Homeopathy and the curse of the scientific method. Veterinary Journal, 180(2), 141-148.

Owen, D., Leckridge, B., & Fisher, P. (2007). Principles and Practice of Homeopathy: The therapeutic and healing process. London: Churchill Livingstone.

Oxman, A. D., Cook, D. J., & Guyatt, G. H. (1994). Users' guides to the medical literature. VI. How to use an overview. Evidence-Based Medicine Working Group. JAMA, 272(17), 1367-1371.

Oxman, A. D., Sackett, D. L., & Guyatt, G. H. (1993). Users' Guide to the Medical Literature 1. How to get started. JAMA, 270(17), 2093-2095.

O'Brien, B. J., Heyland, D., Richardson, W. S., Levine, M., & Drummond, M. F. (1997). Users' guides to the medical literature. XIII. How to use an article on economic analysis of clinical practice. B. What are the results and will they help me in caring for my patients? Evidence-Based Medicine Working Group. JAMA, 277(22), 1802-1806.

O'Dowd, A. (2006). Adverse incidents in NHS are still under-reported. British Medical Journal, 333(7558), 59.

O'Dowd, A. (2009). Giving homoeopathy on the NHS is unethical and unreliable, MPs are told. British Medical Journal, 339, b5080.

PILL Collaborative Group. (2011). An International Randomised Placebo-Controlled Trial of a Four-Component Combination Pill ("Polypill") in People with Raised Cardiovascular Risk. PLoS ONE, 6(5), e19857. doi:10.1371/journal.pone.0019857

Pandolfi, M. (2010). Homeopathy: ex nihilo fit nihil. European Journal of Internal Medicine, 21(3), 147-148.

Pandolfi, M. (2011). Refutable refutations. European Journal of Internal Medicine, 22(1), 118-119.

Papakostas, Y. G., & Daras, M. D. (2001). Placebos, placebo effect, and the response to the healing situation: The evolution of a concept. Epilepsia, 42(12), 1614-1625.

Papineau, D. (1994). The Virtues of Randomization. British Journal for the Philosophy of Science, 45(2), 437 -450.

Park, L. C., & Covi, U. (1965). Nonblind Placebo Trial. Archives of General Psychiatry, 12, 336-345.

Perera, R., & Heneghan, C. (2008). Interpreting meta-analysis in systematic reviews. Evidence Based Medicine, 13(3), 67-69.

Peters, D. (2005). Shang et al. carelessness, collusion, or conspiracy? Journal of Alternative & Complementary Medicine, 11(5), 779–780.

Pilkington, K. (2005). Homeopathy for depression: a systematic review of the research evidence. Homeopathy, 94, 153-163.

Pilkington, K., Kirkwood, G., Rampes, H., Fisher, P., & Richardson, J. (2006). Homeopathy for anxiety and anxiety disorders: a systematic review of the research. Homeopathy, 95(3), 151-162.

Pittrof, R., & Rubenstein, I. (2008). The thinking doctor's guide to placebos. British Medical Journal, 336, 1020-1020.

Pope, C. (2003). Resisting evidence: the study of evidence-based Medicine as a Contemporary Social Movement. Health, 7(3), 267-282.

Popping, R. (2000). Computer-assisted Text Analysis. London: Sage Publications.

Praities, N. (2008a). Homeopathy victim of PCT funding cuts. Pulse News. Retrieved from http://www.pulsetoday.co.uk/newsarticle-content/-/article_display_list/10970138/homeopathy-victim-of-pct-funding-cuts

Praities, N. (2008b). Homeopathic hospital in crisis. Pulse News. Retrieved July 2008, from http://www.pulsetoday.co.uk/story.asp?sectioncode=23&storycode=4120260&c=2

Price, D. D., Finniss, D. G., & Benedetti, F. (2008). A comprehensive review of the placebo effect: recent advances and current thought. Annual Review of Psychology, 59, 565-90.

Price, L. (1984). Art, science, faith and medicine: the implications of the placebo effect. Sociology of Health & Illness, 6(1), 61-74.

Rabkin, J. G., Markowitz, J. S., Stewart, J., McGrath, P., Harrison, W., Quitkin, F. M., & Klein, D. F. (1986). How blind is blind? Assessment of patient and doctor medication guesses in a placebo-controlled trial of imipramine and phenelzine. Psychiatry research, 19(1), 75-86.

Randolph, A. G., Haynes, R Brian, Wyatt, J. C., Cook, D. J., & Guyatt, G. H. (1999). Users' Guides to the Medical Literature: XVIII. How to Use an Article Evaluating the Clinical Impact of a Computer-Based Clinical Decision Support System. JAMA, 282(1), 67-74.

Raoult, D. (2005). Are the clinical effects of homoeopathy placebo effects? – Authors' reply. Lancet, 366, 2085.

Rawlins, M. D. (2008). De Testimonio: on the evidence for decisions about the use of therapeutic interventions. London: Royal College of Physicians of London.

Redelmeier, D. A., & Shumak, S. L. (2003). How to read clinical journals: XI. Everything you always wanted to know about editorials (but were afraid to ask). Canadian Medical Association Journal, 169(12), 1323-1325.

Redelmeier, D. A., Shuchman, M., & Shumak, S. L. (1998). How to read clinical journals: IX. Canadian Medical Association Journal, 159(12), 1488-9.

Relton, C., Smith, C., Raw, J., Walters, C., Adebajo, A. O., Thomas, K. J., & Young, T. A. (2009). Healthcare provided by a homeopath as an adjunct to usual care for Fibromyalgia (FMS): results of a pilot Randomised Controlled Trial. Homeopathy, 98(2), 77-82.

Renckens, C. N. M. (2002). Alternative treatments in reproductive medicine: much ado about nothing. Human Reproduction, 17(3), 528-33.

Richardson, W. S., & Detsky, A. S. (1995a). Users' guides to the medical literature. VII. How to use a clinical decision analysis. B. What are the results and will they help me in caring for my patients?. JAMA, 273(20), 1610-1613.

Richardson, W. S., & Detsky, A. S. (1995b). Users' guides to the medical literature. VII. How to use a clinical decision analysis. A. Are the results of the study valid?. JAMA, 273(16), 1292-1295.

Richardson, W. S., Wilson, M. C., Guyatt, G. H., Cook, D. J., & Nishikawa, J. (1999). Users' Guides to the Medical Literature: XV. How to Use an Article About Disease Probability for Differential Diagnosis. JAMA, 281(13), 1214-1219.

Richardson, W. S., Wilson, M. C., Williams, J. W., Moyer, V. A., & Naylor, C. D. (2000). Users' Guides to the Medical Literature: XXIV. How to Use an Article on the Clinical Manifestations of Disease. JAMA, 284(7), 869-875.

ter Riet, G., de Craen, A. J. M., de Boer, A., & Kessels, A. G. H. (1998). Is placebo analgesia mediated by endogenous opioids? A systematic review. Pain, 76, 273-275.

Riggs, L., Hodgson, S. F., O'Fallon, M., Chao, E. Y. S., Wahner, H. W., Muhs, J. M., Cedel, S. L., et al. (1990). Effect of Fluoride Treatment on the Fracture Rate in Postmenopausal Women with Osteoporosis. New England Journal of Medicine, 322, 802-809.

Robbins, M. (2010a). Quacks fly in all directions as alternative medicine regulation fails. Guardian. Retrieved April 16, 2010, from http://www.guardian.co.uk/science/2010/apr/16/quacks-alternative-medicine-regulation

Robbins, M. (2010b). MPs deliver their damning verdict: Homeopathy is useless and unethical. Guardian. Retrieved February 22, 2010, from http://www.guardian.co.uk/science/blog/2010/feb/22/mps-verdict-homeopathy-useless-unethical

Robbins, M. (2010c). "Choice" fetish spawns homeopathy policy. Guardian. Retrieved July 28, 2010, from http://www.guardian.co.uk/science/2010/jul/27/choice-fetish-homeopathy-policy

Roberts, R. (2010). I don't know how, but homeopathy really does work. Guardian. Retrieved July 15, 2010, from http://www.guardian.co.uk/commentisfree/2010/jul/15/homeopathy-works-scientific-evidence

Rorty, M. V., & Frankel, L. R. (2009). The paradoxical placebo. American Journal of Bioethics, 9(12), 17-20.

Rosenberg, W. M. C., & Donald, A. (1995). Evidence based medicine: an approach to clinical problem solving. British Medical Journal, 310(6987), 1122-1126.

Ross, P. (2010). Homeopathy, a "helpful placebo"; or an unethical intervention? Edzard Ernst Reply. Trends in Pharmacological Sciences, 31(7), 297.

Rossouw, J. E., Anderson, G. L., Prentice, R. L., La Croix, A. Z., Kooperberg, C., Stefanick, M. L., Jackson, R. D., et al. (2002). Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. JAMA, 288(3), 321-333.

Rothstein, W. G. (1992). American Physicians in the Nineteenth Century: from sects to science. London: Johns Hopkins University Press.

Rothwell, P. M. (2005). External validity of randomised controlled trials: "To whom do the results of this trial apply?" Lancet, 365, 82-93.

Roush, S. (2005). Tracking Truth: knowledge, evidence and science. Oxford: Oxford University Press.

Russo, F., & Williamson, J. (2007). Interpreting Causality in the Health Sciences. International Studies in the Philosophy of Science, 21, 1157-1170.

Rutherford, A. (2009). Who's afraid of a homeopath's woo? Guardian. Retrieved October 23, 2009, from http://www.guardian.co.uk/commentisfree/2009/oct/23/religion-medical-research

Rutten, A. L. B., & Stolper, C. F. (2008). The 2005 meta-analysis of homeopathy: the importance of post-publication data. Homeopathy, 97(4), 169-177.

Rutten, A. L. B., & Stolper, C. F. (2009). Reply to Wilson. Homeopathy, 98(2), 129.

Sackett, D. L. (1969). Clinical Epidemiology. American Journal of Epidemiology, 89(2), 125-128.

Sackett, D. L. (2002). Clinical epidemiology. what, who, and whither. Journal of Clinical Epidemiology, 55(12), 1161-1166.

Sackett, D. L., Guyatt, G. H., Tugwell, P, & Haynes, R Brian. (1991). Clinical Epidemiology: A Basic Science for Clinical Medicine (2nd ed.). London: Little, Brown & Company.

Sackett, D. L., Haynes, R Brian, & Tugwell, Peter. (1985). Clinical epidemiology: a basic science for clinical medicine. Toronto: Little, Brown & Company.

Sackett, D. L., Richardson, W. S., Rosenberg, W. M. C., & Haynes, R Brian. (1997). Evidence-based Medicine: How to Practice and Teach EBM. Edinburgh: Churchill-Livingstone.

Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R Brian, & Richardson, W. S. (1996). Evidence based medicine: what it is and what it isn't. British Medical Journal, 312, 71-72.

Sackett, D. L., Straus, S. E., Richardson, W. S., Rosenberg, W. M. C., & Haynes, R Brian. (2000). Evidence-Based Medicine: How to Practice and Teach EBM. Edinburgh: Churchill Livingstone.

Sample, I. (2007). Homeopathy science degrees "gobbledygook." Guardian. Retrieved March 22, 2007, from http://www.guardian.co.uk/science/2007/mar/22/choosingadegree.highereducation

Sample, I. (2008). Boots accused medicines of selling quack medicines. Guardian. Retrieved May 24, 2008, from http://www.guardian.co.uk/science/2008/may/24/controversiesinscience.guardianhayfestival

Sample, I. (2009a). British scientists ask WHO to condemn homeopathy for diseases such as HIV. Guardian. Retrieved June 1, 2009, from http://www.guardian.co.uk/science/2009/jun/01/world-health-organisation-homeopathy-hiv

Sample, I. (2009b). MPs criticise science adviser for defending government on homeopathy. Guardian. Retrieved January 20, 2009, from http://www.guardian.co.uk/science/2009/jan/20/homeopathy-cannabis-john- beddington

Sample, I. (2009c). Homeopathy licences based on "no scientific evidence." Guardian. Retrieved November 21, 2009, from http://www.guardian.co.uk/science/blog/2009/oct/21/pseudoscience

Sample, I. (2010a). NHS funding for homeopathy risks misleading patients, says chief scientist. Guardian. Retrieved October 27, 2010, from http://www.guardian.co.uk/science/2010/oct/27/nhs-funding-homeopathy-chief-scientist

Sample, I. (2010b). Homeopathic society "misled" MPs in inquiry. Guardian. Retrieved February 5, 2010, from http://www.guardian.co.uk/science/2010/feb/05/homeopathy-false-evidence-parliamentary-inquiry

Schapira, K., McClelland, H. A., Griffiths, N. R., & Newell, D. J. (1970). Study on the effects of tablet colour in the treatment of anxiety states. British Medical Journal, 2(5707), 446.

Scheper-Hughes, N., & Lock, M. M. (1987). The Mindful Body: A Prolegomenon to Future Work in Medical Anthropology. Medical Anthropology Quarterly, 1(1), 6-41.

Schwab, A. P. (2009). When subtle deception turns into an outright lie. American Journal of Bioethics, 9(12), 30-32.

Science Museum. (2011). Living Medical Traditions. Retrieved February 14, 2011, from http://sciencemuseumdiscovery.com/blogs/collections/living-medical-traditions/

Sehon, S., & Stanley, D. (2010). Evidence and simplicity: why we should reject homeopathy. Journal of Evaluation in Clinical Practice, 16(2), 276-281.

Sense About Science. (2006). Sense about Homeopathy. <http://www.senseaboutscience.org/data/files/resources/54/Homeopathy.pdf>

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and Quasi-Experimental Designs for Generalized Causal Inference. New York: Houghton Mifflin Company.

Shang, A., Huwiler-Müntener, K., Nartey, L., Jüni, P., Dörig, S., Sterne, J. A., Pewsner, D., et al. (2005). Are the clinical effects of homoeopathy placebo effects? Comparative study of placebo-controlled trials of homoeopathy and allopathy. Lancet, 366(9487), 726-732.

Shang, A., Jüni, P., Sterne, J. A., Huwiler-Müntener, K., & Egger, M. (2005). Are the clinical effects of homoeopathy placebo effects? – Authors' reply. Lancet, 366(9503), 2083-2085.

Shapiro, A. K. (1964). historic and heuristic definition of the placebo'. Psychiatry, 77(1), 52-8.

Shapiro, A. K. (1968). Semantics of the placebo. Psychiatric Quarterly, 42(4), 653-695.

Shapiro, A. K., & Morris, L. A. (1978). The placebo effect in medical and psychological therapies. In S. L. Garfield & A. E. Bergin (Eds.), Handbook of Psychotherapy and Behavior Change (Second Edition). New York: Wiley.

Shapiro, A. K., & Shapiro, E. (1997a). The Powerful Placebo. London: Johns Hopkins University Press.

Shapiro, A. K., & Shapiro, E. (1997b). The Placebo: Is It Much Ado About Nothing? In Anne Harrington (Ed.), The Placebo Effect: An Interdisciplinary Exploration. London: Harvard University Press.

Shapiro, A. K., & Shapiro, E. (1999). The powerful placebo: From ancient priest to modern physician. Bulletin of the History of Medicine, 73, 159-160.

Shapiro, R. (2008). Suckers: how alternative medicine makes fools of us all. London: Harvill Secker.

Shumak, S. L., & Redelmeier, D. A. (2000). How to read clinical journals: X. How to react when your colleagues haven't read a thing. Canadian Medical Association Journal, 163(12), 1570-1572.

Shumak, S. L., & Redelmeier, D. A. (2004). How to read clinical journals: XII. How you too can profit from pharmaceutical advertisements. Canadian Medical Association Journal, 171(12), 1455-1456.

Silverman, W. A. (2004). A cautionary tale about supplemental oxygen: the albatross of neonatal medicine. Pediatrics, 113(2), 394-396.

Simpson, S. J. Y. (1853). Homoeopathy: its tenets and tendencies, theoretical, theological, and therapeutical. London: Sutherland and Knox.

Singh, S. (2009). Tiddles the cat enjoys better protection from charlatans than you or I. Guardian. Retrieved April 24, 2009, from http://www.guardian.co.uk/science/2009/apr/24/homeopathy-pets-vets-animals-placebo

Singh, S., & Ernst, Edzard. (2009). Trick Or Treatment?: Alternative Medicine on Trial (p. 416). London: Bantam Press.

Skandhan, K. P., Amith, S., & Avni, S. (2005). Are the clinical effects of homoeopathy placebo effects? – Authors' reply. Lancet, 366, 2085.

Skinner, S. (2001). An Introduction to Homeopathic Medicine in Primary Care. Maryland, USA: Aspen Publishers.

Smith, G. C. S., & Pell, J. P. (2003). Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. British Medical Journal, 327(7429), 1459-1461.

Smith, K. (2011). Against Homeopathy - a Utilitarian Perspective [EARLY VIEW]. Bioethics, 1-12. doi:10.1111/j.1467-8519.2010.01876.x

Society Of Homeopaths. (2010). The Society of Homeopaths condemns 10.23 campaign "overdose" as an ill advised publicity stunt in poor taste. Retrieved February 2010, from http://www.homeopathy-soh.org/whats-new/past-press-releases/documents/BBCStatementAug09.doc

Stanley, D., & Sehon, S. (2011). Taking Procrustes' Axe to Professor Fisher's response. Journal of Evaluation in Clinical Practice, 17(5), 1009-1010.

Stanley, J., & Williamson, T. (2001). Knowing How. Journal of Philosophy, 98(8), 411-444.

Stein, H. F. (1983). . .. On Placebos. To Cure, To Control, To Please: Medicine after the Demise of "The Placebo." Medical Anthropology Quarterly, 15(1), 4-17.

Stewart-williams, S., & Podd, J. (2004). The placebo effect: Dissolving the expectancy versus conditioning debate. Psychological Bulletin, 130(2), 324-340.

Straus, S. E., & McAlister, F. A. (2000). Evidence-based medicine: a commentary on common criticisms. Canadian Medical Association Journal, 163(7), 837-841.

Straus, S. E., Haynes, R Brian, Glasziou, P., Dickersin, K., & Guyatt, G. H. (2007). Misunderstandings, misperceptions, and mistakes. Evidence Based Medicine, 12, 2-3.

Straus, S. E., Richardson, W. S., Glasziou, P., & Haynes, R Brian. (2005). Evidence-Based Medicine: How to Practice and Teach EBM (Third Edition). London: Elsevier.

Streitberger, K., & Kleinhenz, J. (1998). Introducing a placebo needle into acupuncture research. Lancet, 352(9125), 364-365.

Sullivan, M. (1993). Placebo controls and epistemic controls in orthodox medicine. Journal of Medicine and Philosophy, 18(2), 213-231.

Tanenbaum, S. J. (1993). What Physicians Know. New England Journal of Medicine, 329(17), 1268-1271.

Taylor, D. W., Barnett, H. J., Haynes, R B, Ferguson, G. G., Sackett, D. L., Thorpe, K. E., Simard, D., et al. (1999). Low-dose and high-dose acetylsalicylic acid for patients undergoing carotid endarterectomy: a randomised controlled trial. Lancet, 353, 2179-84.

Temple, R., & Ellenberg, S. S. (2000). Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 1: ethical and scientific issues. Annals of Internal Medicine, 133(6), 455-463.

The Cardiac Arrhythmia Suppression Trial (CAST) Investigators. (1989). Preliminary report: effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. New England Journal of Medicine, 321(6), 406-412.

The EC/IC Bypass Study Group. (1985). Failure of extracranial-intracranial arterial bypass to reduce the risk of ischemic stroke. Results of an international randomized trial. New England Journal of Medicine, 313(19), 1191-1200.

Thompson, J. J., Ritenbaugh, C., & Nichter, M. (2009). Reconsidering the Placebo Response from a Broad Anthropological Perspective. Culture, Medicine and Psychiatry, 33, 112-152.

Thompson, T. D. B., & Weiss, M. (2006). Homeopathy--what are the active ingredients? An exploratory study using the UK Medical Research Council's framework for the evaluation of complex interventions. BMC Complementary and Alternative Medicine, 6:37, 1-16.

Tonelli, M. R. (1998). The Philosophical Limits of Evidence-based Medicine. Academic Medicine, 73(12), 1234-1240.

Turner, A. (2012). "Placebos" and the logic of placebo comparison. Biology & Philosophy, 27(3), 419-432.

U.S. National Library of Medicine. (2009). Homeopathy - MeSH Descriptor Data. http://www.nlm.nih.gov/cgi/mesh/2009/MB_cgi: 18/09/2009.

Upshur, R. E. G., & Colak, E. (2003). Argumentation and evidence. Theoretical Medicine and Bioethics, 24(4), 283-299.

Vase, L., Nørskov, K. N., Petersen, G. L., & Price, D. D. (2011). Patients' direct experiences as central elements of placebo analgesia. Philosophical Transactions of the Royal Society B, 366(1572), 1913-1921.

Vickers, A. J. (2000). Clinical trials of homeopathy and placebo: Analysis of a scientific debate. Journal of Alternative & Complementary Medicine, 6(1), 49–56.

Vickers, A. J., & Zollman, C. (1999). ABC of complementary medicine: Homoeopathy. British Medical Journal, 319, 1115-1118.

Wahlberg, A. (2008). Above and beyond superstition - western herbal medicine and the decriminalizing of placebo. History of the Human Sciences, 21(1), 77-101.

Walach, H. (2001). The efficacy paradox in randomized controlled trials of CAM and elsewhere: beware of the placebo trap. Journal of Alternative & Complementary Medicine, 7(3), 213-218.

Walach, H. (2003). Entanglement model of homeopathy as an example of generalized entanglement predicted by weak quantum theory. Forsch Komplementarmed Klass Naturheilkd, 10(4), 192–200.

Walach, H. (2005). Generalized entanglement: A new theoretical model for understanding the effects of complementary and alternative medicine. Journal of Alternative & Complementary Medicine, 11(3), 549–559.

Walach, H. (2011). Placebo controls: historical, methodological and general aspects. Philosophical Transactions of the Royal Society B, 366(1572), 1870-1878.

Walach, H., Jonas, W. B., & Lewith, G. T. (2005). Are the clinical effects of homoeopathy placebo effects? Lancet, 366, 2081.

Walsh, B. T., Seidman, S. N., Sysko, R., & Gould, M. (2002). Placebo Response in Studies of Major Depression: Variable, Substantial, and Growing. JAMA, 287(14), 1840-1847.

Van Wassenhoven, M., & Ives, G. (2004). An observational study of patients receiving homeopathic treatment. Homeopathy, 93, 3-11.

Weatherley-Jones, E., Thompson, E. A., & Thomas, K. J. (2004). The placebo-controlled trial as a test of complementary and alternative medicine: observations from research experience of individualised homeopathic treatment. Homeopathy, 93(4), 186-189.

Weber, R. P. (1984). Computer Aided Content Analysis: A short primer. Qualitative Sociology, 7(1-2), 126-147.

Weingärtner, O. (2007). The nature of the active ingredient in ultramolecular dilutions. Homeopathy, 96(3), 220-226.

Weissman, J. S., Blumenthal, D., Silk, A. J., Zapert, K., Newman, M., & Leitman, R. (2003). Consumers' reports on the health effects of direct-to-consumer drug advertising. Health Affairs, 3, 82-95.

West Kent Primary Care Trust. (2007a). Homeopathy Commissioning Review. <http://www.westkentpct.nhs.uk/>

West Kent Primary Care Trust. (2007b). Should the NHS pay for homeopathy? SUMMARY OF WEST KENT PCT'S PUBLIC CONSULTATION PROCESS AND FEEDBACK . <http://www.westkentpct.nhs.uk/>

West Kent Primary Care Trust. (2009). Homeopathy Commissioning Review Final Report. <http://www.westkentpct.nhs.uk/>

West, S., King, V., Carey, T. S., Lohr, K. N., McKoy, N., Sutton, S. F., & Lux, L. (2002). Systems to rate the strength of scientific evidence. Rockville, MD. <http://www.thecre.com/pdf/ahrq-system-strength.pdf>

White, C. (2010). Local bodies should decide whether to fund homoeopathy, says government. British Medical Journal, 341, c4073.

White, P., Lewith, G. T., Hopwood, V., & Prescott, P. (2003). The placebo needle, is it a valid and convincing placebo for use in acupuncture trials? A randomised, single-blind, cross-over pilot trial. Pain, 106(3), 401-409.

Williams, B. A. (2010). Perils of evidence-based medicine. Perspectives in Biology and Medicine, 53(1), 106-120.

Williamson, T. (1992). Inexact Knowledge. Mind, 101(402), 217-242.

Williamson, T. (2000). Knowledge and its Limits. Oxford: Oxford University Press.

Williamson, T. (2007). The Philosophy of Philosophy. Oxford: Blackwell.

Wilson, K. (2010). Evidence-based medicine. The good the bad and the ugly. A clinician's perspective. Journal of Evaluation in Clinical Practice, 16(2), 398-400.

Wilson, M. C., Hayward, R. S., Tunis, S. R., Bass, E. B., & Guyatt, G. H. (1995). Users' guides to the Medical Literature. VIII. How to use clinical practice guidelines. B. what are the recommendations and will they help you in caring for your patients? The Evidence-Based Medicine Working Group. JAMA, 274(20), 1630-1632.

Wilson, Paul. (2009). Analysis of a re-analysis of a meta-analysis: in defence of Shang et al. Homeopathy, 98, 127-128.

Winfield, A. J., & Kennedy, E. J. (2004). Oral Unit Dosage Forms. In A. J. Winfield & R. M. E. Richards (Eds.), Pharmaceutical Practice (3rd ed.). Edinburgh: Churchill Livingstone.

Worrall, J. (2002). What Evidence in Evidence-Based Medicine? Philosophy of Science, 69(3), S316-S330.

Worrall, J. (2007a). Why there's no cause to randomize. British Journal for the Philosophy of Science, 58, 451-488.

Worrall, J. (2007b). Evidence in medicine and evidence-based medicine. Philosophy Compass, 2(6), 981-1022.

Worrall, J. (2008). Evidence and ethics in medicine. Perspectives in Biology and Medicine, 51(3), 418-31.

Worrall, J. (2010). Evidence: philosophy of science meets medicine. Journal of Evaluation in Clinical Practice, 16(2), 356-362.

Yu-Hin Ng, D. (2011). A discussion: the future role of homeopathy in the National Health Service (NHS). Homeopathy, 100(3), 183-186.

Zimmermann-Viehoff, F., & Meissner, K. (2007). Homeopathy and Placebo – Synonym, Similar or Different? Forschende Komplementärmedizin, 14(4), 247-248.