CHAPTER 16

Reading 16.1 **Geddes, J.R. and Harrison, P.J. (1997). Closing the gap between research and practice. *British Journal of Psychiatry*, 171: 220–225.**

Reading 16.2 **Hume, D. ([1748] 1975). *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*. Oxford: Oxford University Press, section IV (Extract pp. 37–8).**

Reading 16.3 **Popper, K. (1972). Conjectural knowledge: my solution to the problem of induction. In *Objective Knowledge*. Oxford: Oxford University Press, pp. 1–31 (Extract pp. 7–9).**

Reading 16.4 **Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In *Criticism and the Growth of Knowledge* (ed. I. Lakatos and A. Musgrave). Cambridge: Cambridge University Press, pp. 91–195 (Extract pp. 100–101).**

Reading 16.5 **Kuhn, T. S. (1970). Logic or discovery or psychology of research? In *Criticism and the Growth of Knowledge* (ed. I. Lakatos and A. Musgrave). Cambridge: Cambridge University Press, pp. 1–23 (Extract pp. 4–6).**

Reading 16.6 **Bloor, D. (1976). *Knowledge and Social Imagery.* London: Routledge, (Extract pp. 1–3).**

Reading 16.7 **Mellor, D.H. (1991). The warrant of induction. In *Matters of Metaphysics*. London: Routledge. (Extract sections 1 and 6).**

Reading 16.8 **Wittgenstein, L. (1969). *On Certainty*, Oxford: Basil Blackwell (Extract various paragraphs)**

Readings 16.1

EXERCISE 1

From: Geddes, J.R. and Harrison, P.J. (1997). Closing the gap between research and practice. *British Journal of Psychiatry*, 171: 220–225.

Has Mrs A got schizophrenia? Would Mr B do better with a tricyclic or a selective serotonin reuptake inhibitor? Is Miss C likely to be recovered in six months time? Is St John's wort an effective antidepressant?

Clinical psychiatry involves making difficult decisions about diagnosis, therapy and prognosis. Sometimes we may be entirely confident about our decisions, but often we are uncomfortably aware that we are making a choice without being sure there is convincing evidence to justify it. Maybe we don't know or have forgotten what the evidence is, or perhaps there isn't any.

All doctors have considerable information needs which are often unrecognised and unmet (Smith, 1996). Clinicians turn to colleagues, textbooks and reviews to keep abreast of developments and to inform their decision-making. However these sources suffer from a number of limitations and biases—and may be out of date. For example, conventional review articles may appear authoritative but contain a considerable, and often unknown, degree of subjective opinion, such that "deciding which review to believe is like deciding which toothpaste to use. It is a question of taste rather than a question of science" (Oxman & Guyatt, 1988). On the other hand, it is unrealistic to expect anyone to locate, let alone keep up with, the two million papers published in 20 000 biomedical journal every year (Murlow, 1994) or to identify the tiny fraction which are both scientifically sound and clinically important to psychiatrists.

The effect of this unmet need for accurate information is a 'knowledge gap' whereby many clinical decisions are made with a greater degree of opinion and uncertainty than is necessary. The gap is filled by a number of other factors which together influence the decisions we make: such factors include the conceptual aetiological school to which we subscribe (e.g. biological *v.* social psychiatry) and the combination of experience and habits which we accumulate during our career. In this context, it is easy to understand the marked differences between psychiatrists' decisions, for example in the use of electroconvulsive therapy (Pippard, 1992; Hermann *et al*, 1995), continuation antipsychotics (Meise *et al*, 1994) and the treatment of depression (Wells *et al*, 1994). Variation in therapeutic (and diagnostic) practice is justified, even desirable, if all the variants are equally effective or if there really is no evidence. However, most psychiatrists would agree this is unlikely always to be the case, and that better application of the existing evidence would lead to greater uniformity and to higher overall standards of psychiatric care with improved outcomes.

Given the knowledge gap and reason to doubt the effectiveness and objectivity of traditional ways of keeping up with developments in clinical practice, psychiatrists need a new strategy to ensure that their clinical decision-making is based upon the most appropriate, accurate, and up-to-date information: evidence-based medicine (EBM).

What is Evidence-Based Medicine?

EBM is the "conscientious, explicit and judicious use of the current best evidence in making decisions about the care of individual patients" (Sackett *et al*, 1996*a*). EBM closes the gap between research and practice by incorporating the advances in clinical epidemiology and medical informatics into clinical activities (Evidence-Based Medicine Working Group, 1992). EBM is intended to overcome the problems outlined above: that clinicians may not always make the best decisions; that they may not realize this and, if they do, they may not realize this and, if they do, they may not know how to improve the situation. By providing clinicians with a set of skills which allow them to base clinical decisions on the best available, most up-to-date evidence, EBM also aims to be a method of self-directed, career-long learning.

Views about EBM tend to be polarized. Critics of EBM caricature its proponents as evangelists who fail to appreciate the complexity of everyday medical practice and who overlook the wisdom of experienced clinicians; the advocates of EBM view its opponents as Luddites who have an overvalued opinion of their clinical acumen. In this paper we have tried to address both concerns. We outline the principles and practice of EBM as it applies to psychiatry. We focus upon the use of EBM to help a psychiatrist to make his or routine clinical decisions more efficient and evidence-based. We also recognize and discuss the various limitations of EBM. Our stance, however, is unequivocally in favour of EBM; the only sensible debate concerns how it can be applied most effectively and pragmatically.

How to practise Evidence-Based Medicine

The practice of EBM focuses on five linked activities, as outlined below.

Defining the clinical question to be answered

Clear answers require clear questions. The formulation of a precise clinical question involves definition of the patient's problem, identification of the manoeuvre to be performed, and specification of the outcome of interest.

Example
Encouraged by the health authority, your Trust is trying to improve the local implementation of the Care Programme Approach (CPA). The official guidance, which introduces the concept of 'tiered' CPA for different levels of clinical need, is open to several interpretations. Several of your colleagues suggest that the CPA is worthless, that resources for its implementation are not available, and that the government should be encouraged to

withdraw it. One cites a damning editorial (Anonymous, 1995). Other colleagues feel that it should be fully implemented for all patients and that the purchasers should be billed accordingly. You suggests that although the use of the CPA is compulsory, it will be easier to implement if you know what effect it might be expected to have. You therefore decide to look for some evidence on its effectiveness. You pose the question: "For patients with schizophrenia or other severe mental illness, what are the effects of CPA on the patient's clinical state and service utilization?"

Searching for the evidence

Recent advances in information technology have made it far easier to search the medical literature. Access to electronic databases such as the National Library of Medicine's Medline is now widespread in medical libraries, and the British Medical Association offers free Medline access to members with computers and modems. Increasingly, medical journals offer World Wide Web pages on the Internet, displaying contents, abstracts and some full text articles.

Electronic databases are useful but they have limitations. Not all journals are indexed, the relevant articles can be difficult to locate, and the searcher's access and time may be limited. One solution to these problems is to use systematic reviews prepared by others. These are a major advance. Unlike a conventional reviewer, a systematic reviewer uses explicit methods of searching for and critically appraising the primary studies. If these are comparable, the reviewer may then perform a formal quantitative synthesis (meta-analysis) of the results (Mulrow, 1987). As well as summarizing the effectiveness of health care interventions, a systematic review can reveal problems with the literature, for example revealing duplicate publications (Huston & Moher, 1996). Two increasingly useful sources of systematic reviews of health care interventions, including psychiatric treatments, are the Cochrane collaboration (1996) and the NHS Centre for Reviews and Dissemination (Sheldon, 1996).

While these initiatives fulfil an essential role, there is inevitably a delay between important new evidence appearing and it being incorporated into a systematic review. There is therefore a need for a method of identifying, critically appraising and disseminating new evidence as it becomes available. One solution has been the introduction of two journals of secondary publication—the *American College of Physicians (ACP) Journal Club* and *Evidence-Based Medicine* (which both include psychiatric topics). Their aim is to locate and summarise the small number of articles, concerning advances in diagnosis, therapy, prognosis, aetiology and economic evaluation which are clinically important, and to publish them in the form of structured abstracts prepared by trained clinical epidemiologists with commentaries provided by experienced clinicians. Now a new journal—*Evidence-Based Mental Health*—is being planned to meet the psychiatrist's needs more specifically.

Example contd:

You try to find a relevant systematic review of randomized controlled trials (RCTs). CPA is really a variant of case management and so you search for reviews of 'case management' in the Cochrane Database of Systematic Reviews on the Cochrane Library. You find a systematic review of case management for people with severe mental disorders which appears to meet your needs (Marshall *et al*, 1996).

Appraising the evidence

Evidence needs to be critically appraised for its reliability and whether it can be extrapolated to the particular clinical problem. A fundamental principle of EBM is that evidence from clinical research studies is more trustworthy than reasoning from basic pathophysiology, psychology or clinical intuition (Evidence-Based Medicine Working Group, 1992). Implicit in the process of critical appraisal is the idea that there is critical appraisal is the idea that there is hierarchy in the quality of evidence derived from different types of study, which is inversely proportional to their susceptibility to bias (that is—the systematic deviation of the results from the truth). A commonly used hierarchy of research designs for evaluating therapy is shown in Table 1.

RCTs, or better still, systematic reviews of RCTs, are the most reliable study design for the evaluation of treatments. However, for many interventions, RCTs may not exist and the practitioner has to use evidence from the next level of the hierarchy. The practical importance of the hierarchy is that, when choosing between treatments, the clinical selects the intervention which is backed up by the best available evidence. When evaluating a paper evaluating therapy, the main questions to ask are those addressing the study's internal validity (are the results trustworthy?) and external validity (can they be applied to my patients?) (Table 2) (Sackett *et al*, 1997).

Table 1 A hierarchy of evidence for therapy

Ia.	Evidence from a meta-analysis of RCTs
Ib.	Evidence from at least one RCT
IIa.	Evidence from at least one controlled study without randomisation
IIb.	Evidence from at least one other type of quasi-experimental study
III.	Evidence from non-experimental descriptive studies, such as comparative studies, correlation studies and case-control studies
IV.	Evidence from expert committee reports, or opinions and/or clinical experience of respected authorities

Table 2 Questions for evaluating an article about therapy

Are the results of this trial valid?
Was the assignment of patients to treatments truly randomised?
Were the groups similar at the start of the trial?
Were patients and clinicians kept blind to which treament was being received?
Aside from the experimental treatment, were the groups treated equally?
Were all the patients who entered the trial accounted for at its conclusion?
The evidence is valid and important. Should it be applied to my patient?
Is the patient so different from those in the trial that the results can't help me?
How great would the potential benefit of therapy actually be for my patient?
Are my patient's values and preferences satisfied by the regimen and its consequences?

Table 3 Guidelines for evaluating an overview of randomized controlled trials

Is it an overview of RCTs of the treatment you're interested in?
Does it include a methods section that describes: Whether and how all the relevant trials were identified and included? How the validity of the individual trials was assessed?
Were the results consistent from study to study?

Critical appraisal of an overview of RCTs asks several further questions (Table 3).

Example contd:
Your critically appraise the Cochrane review for its validity and applicability to your problem. The review includes nine RCTs of case management *v.* standard care. The search strategy and inclusion and exclusion criteria are clearly described and there is an assessment of the quality of the individual studies. All used a form of case management which was comparable to that included in the CPA: standard care was defined as the normal level of psychiatric care provided in the area where the trial was conducted. The studies examined a range of outcomes including: loss to follow-up, admission to hospital, death, mental state at one year, social functioning at one year, quality of life at one year and health care costs. There was some variation in results from study to study, but the reasons for this are investigated and the variation is mainly quantitative rather than qualitative. You decide that you can believe the results of the overview. The studies in the overview were performed in the UK and the USA and included patients with severe mental illness (however defined). This is likely to include the group of patients which interest you.

Applying the evidence

Translating research findings into clinically usable information is one of the most challenging areas of EBM. It involves summarizing the research findings in ways which contain the maximum amount of information and which are meaningful to the clinician (or to patients or purchasers). Many studies present their results in terms of ratio measures such as the odds ratio and the relative risk (see Box 1). However this can lead to overestimates of the treatment effect because they do not give any idea of the control event rate (Fahey *et al*, 1995; Sackett *et al*, 1996*b*). For example, if a treatment reduces the risk of death, its effect would be less impressive if the baseline mortality was 2% (reduced to 1% by the treatment) than if the baseline risk was 80% (reduced to 40% by the treatment), although in both instances the relative risk of death is 0.5. Most commentators favour measures of effect such as the absolute risk reduction or the number needed to treat, which are more intuitive and more clinically useful (Laupacis *et al*, 1998). Unfortunately at present, the reader often has to calculate these indices from the review.

Patients in RCTs are likely to be different from the majority of patients encountered in real-life clinical practice, for example because patients consenting to enter the trial are unrepresentative

Box 1 Indices for translating research results into clinical practice

	Experimental treatment, X	*Control treatment, Y*
Positive outcome	*a*	*b*
Negative outcome	*c*	*d*

Control Event Rate (*CER*) = b/(b+d)
Experimental Event Rate (*EER*) = a/(a+c)
Absolve Risk Reduction (*ARR*)
The difference in the proportions with a positive outcome on treatments X and Y (=CER−EER)
Relative risk (*RR*)
The ratio of the proportions with a positive outcome on treatments X and Y = [a/(a+c)]/[b/b+d)] =EER/CER
Odds ratio (*OR*)
The ratio of the odds of a positive outcome on treatments X and Y (= (a/c)/(b/d) =ad/bc)
Number Needed to Treat (NNT)—how many patients need to be treated with treatment X to get one more positive outcome than would be expected on treatment Y (=1/ARR)

of all subjects approached, or because comorbidity is an exclusion criterion. Hence the need for pragmatic RCTs on representative patients (Simon *et al,* 1995), and for extrapolation of RCT results to clinical situations by combining them with appropriate event rates derived from cohort studies, or by the use of local data (see below: Cook & Sackett, 1995; Glasziou & Irwig, 1995).

Example contd:
The main findings of the review were that case management assists teams to maintain contact with patients but also increases the admission rate. There was no significant effect on the other outcome variables (such as reduction in clinical symptoms).

Loss to follow-up—151 of 600 (Experimental Event Rate (EER)=25.2%)case managed subjects were lost to follow-up compared with 200 of 611 (Control Event Rate (CER)=32.7%) standard care subjects. The Absolute Risk Reduction (ARR) is 7.5% and the Number Needed to Treat (NNT) is 13 (95% CI 8–40). This means that 13 patients have to be treated with case management to prevent one less patient from being lost to follow-up than would occur with standard care. The ratio of the odds of being lost to follow-up in the case management group to the odds of being lost to follow-up in the standard care group was 0.68 (95% CI 0.53–0.88). You know from an audit of your service, carried out before the introduction of the CPA, that only 10% of patients with severe mental illness were lost to follow-up each year. This is about 30% (0.100/0.327=0.306) of the CER in the review. Assuming that case management has a constant effect in all patient groups, you can therefore adjust the NNT to apply to the loss to follow-up rate of your service by dividing it by this proportion (Cook & Sackett, 1995). This produces a NNT of about 42 (13/0.306). However, among patients with a previous history of loss of contact, your drop-out rate was 80%. Adjustment in a similar way produces an NNT estimate of about five for this group of patients.

> *Admission to hospital*—One of the key aims of community care is to reduce the rate of hospital admission. In the review, case management actually *increased* the risk of admission; the CER was 19.4 and the EER 30.7. The ARR was—11.3% and number needed to harm (NNH) was 9 (6–15). This means that for every nine patients case managed one more would be admitted to hospital than would have occurred if standard care had been used. Again this figure could be adjusted to fit local readmission rates by making the same assumption of constant effect across subgroups.
>
> You present these figures to your working group. The group concludes that CPA is likely to be most useful for keeping contact with patients who have previously dropped out of care. They will concentrate on implementing 'full' CPA for this group initially, expanding to other groups as resources become available. The in-patient service will need to be supported to cope with the predicted increased demand. (The group also noted that admission could be viewed as a positive event rather than as a negative one—it is decided to investigate reason for admissions locally.)

The example makes it clear that interpretation of evidence, and health care decisions, are complex even when good quality evidence is available and when EBM is adopted. Evidence is helpful but not sufficient for medical decision-making: the key aspect of EBM is that it ensures the best use is made of the *available* evidence.

We have used an example of an evidence-based approach to a decision about therapy. The same EBM process can be applied to decisions about diagnosis (e.g. how useful are screening instruments for detecting cases of depression in primary care? How useful are Schneider's first-rank symptoms in diagnosing schizophrenia?) (Geddes *et al*, 1996*a*) or prognosis (e.g. what is the risk of suicide in someone who has taken an overdose? What is the outcome following a first episode of schizophrenia?). The Evidence-Based Medicine Working Group is publishing a series of articles in the *Journal of the American Medical Association* which provide a guide to how to appraise and interpret articles when using EBM to answer these different types of question (Table 4).

Evaluation and development of evidence-based practice

The final stage of EBM involves the development and evaluation of one's own EBM skills: the ability to ask answerable questions, to search for evidence, and to appraise it and to apply it appropriately. Through a recursive process, gaps in practice or teaching can be identified and rectified, and the proportion of one's clinical decisions which are soundly evidence-based can be audited.

There are several ways of developing the numerical and critical appraisal skills necessary for EBM. EBM teaching workshops are becoming widely available. 'Educational prescriptions' are useful for introducing EBM into teaching and into one's own clinical practice (Sackett *et al*, 1997). A prescription defines a problem for which uncertainity exists and specifies a time and date for the question to be answered. Evidence is then sought, critically appraised, and applied to the problem. Academic meetings and journal clubs can be similarly structured (Gilbody, 1996). The content of continuing medical education, especially now that it is becoming obligatory, should obviously be evidence-based too, and effective strategies used to deliver it (Davis *et al*, 1995). Finally, practice in general can become more evidence-based by the adoption of evidence-based guidelines, often devised nationally, but critically appraised and tailored for local implementation (NHS Center for Reviews and Dissemination, 1994; Eccles *et al*, 1996).

Table 4 The Evidence-Based Medicine Working Group's "Users guide to the medical literature" series published in the *Journal of the American Medical Association* (*JAMA*)

Oxman, A. D., Sackett, D. L. & Guyaat, G. H. (1993) I. How to get started. *JAMA*, **270**, 2093–2095.

Guyatt, G. H., Sackett, D. L. & Cook, D. J. (1993) II. How to use an article about therapy or prevention. A. Are the results of the study valid? *JAMA*, **270**, 2598–2601.

—,—&— (1994) 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.

Jaeschke, R., Guyatt, G. H. & Sackett, D. L. (1994) III. How to use an article about a diagnostic test. A. Are the results of the study valid? *JAMA*, **271,** 389–391.

—,—&— (1994) 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? *JAMA*, **271**, 703–707.

Levine, M., Walter, S., Lee, H., *et al* (1994) IV. How to use an article about harm. *JAMA*, **271**, 1615–1619.

Laupacis, A., Wells, G., Richardson, W. S., *et al* (1994) V. How to use an article about prognosis. *JAMA*, **272**, 234–237.

Oxman, A. D., Cook, D. J. & Guyatt, G. H. (1994) VI. How to use an overview. *JAMA*, **272**, 1367–1371.

Richardson, W. S. & Detsky, A. S. (1995) VIII. How to use a clinical decision analysis. A. Are the results of the study valid? *JAMA,* **273**, 1292–1295.

Hayward, R. S., Wilson, M. C., Tunis, S. R., *et al* (1995) VIII. How to use clinical practice guidelines. A. Are the recommendations valid? *JAMA*, **274**, 570–574.

Wilson, M. C., Hayward, R. S., Tunis, S. R., *et al* (1995) VIII. How to use clinical practice guidelines. B. What are the recommendations and will they help you in caring for your patients? *JAMA*, **274**, 1630–1632.

Guyatt, G. H., Sackett, D. L., Sinclair, J. C., *et al* (1995) IX. A method for grading health care recommendations. *JAMA*, **274,** 1800–1804.

Naylor, C. D. & Guyatt, G. H. (1996) X. How to use an article reporting variations in the outcomes of health services. *JAMA*, **275**, 554–558.

Addressing concerns about Evidence-Based Medicine

Apart from *ad hominem* attacks (Shahar, 1995), EBM has been criticized on a number of grounds. These have often arisen from a misunderstanding of what EBM is and what it can do (Sackett *et al*, 1997). However, there remain some important concerns about the theoretical and practical implications of EBM.

Is EBM really different from what good clinicians have always practised?

The idea that practice should be based on good evidence is hardly new, and the aims of EBM are the same as traditional medical practice (Lewis, 1958). While exceptional clinicians may always have kept fully up-to-date without using EBM, the existence of variations in practice implies that this is not generally true for all clinicians. At the very least, by making the link between evidence and practice efficient and explicit, EBM allows outstanding clinicians to demonstrate their prowess —and allows the rest of us to emulate them.

Hypocrisy: the strategies of EBM are flawed, and there is no evidence that EBM works

The methods of EBM are still being developed and they need to be used carefully and critically (Sackett *et al*, 1996*a*). For example, it is known that meta-analysis can be misleading (Eysenck, 1994; Egger & Smith, 1995), and that interpretation of the same evidence can come up with different conclusions (Jackson & Sackett, 1996). Such factors point to the need for continuing refinement and testing of the EBM tools, including meta-analysis (Sim & Hlatky, 1996), and emphasise that the clinician must continue to use his or her critical faculties when appraising evidence, even when it is presented on an evidence-based plate in the form of a systematic review. EBM does not claim to be an infallible, discrete system which renders other approaches invalid; rather, it complements and strengthens traditional medical skills.

A related claim of hypocrisy is that there is no proof that EBM is effective. Ideally, this would require an RCT showing that the practice of EBM improves patient outcome. Though there is no evidence for the effectiveness of EBM as a system-level intervention (and one could argue that such a study is not feasible), there are empirical findings suggesting that the elements of EBM do achieve their aims. First, the effectiveness of critical appraisal (essentially, the adoption of an EBM approach to medicine) has been shown; medical students trained in this way were more likely than their peers to use critical appraisal skills and keep up-to-date with therapeutic advances several years after qualifying (Bemmett *et al*, 1987; Shin *et al*, 1993). Second, recent observations suggest that the application of EBM strategies improves outcome—for example the fall in breast cancer death rates since the publication of the meta-analysis by the Early Breast Cancer Trialist Collaborative Group (Beral *et al*, 1995).

We would also argue that EBM has high face validity. EBM represents an explicit, coherent and comprehensive approach, enabling clinicians to ensure that their practice is based on the best available evidence. Although formally unproven, it is difficult to challenge the assumption that this will lead to the best clinical outcomes.

Evidence is lacking, and what there is is inapplicable to real-life practice

The proportion of medical interventions which are bases on evidence has been underestimated (Smith, 1991; Ellis *et al*, 1995). In psychiatry, the tendency may be more marked because of the widely-held beliefs about the nature of mental illness (Kerr *et al*, 1995; Priest *et al*, 1996). This can lead to pessimism about the feasibility of applying EBM to psychiatry because it may be thought that the speciality comprises islands of evidence in a sea of opinion, and that the interventions for which evidence does exist are self-evidently effective anyway. In fact, the proportion of treatment decisions in psychiatry which are based on RCT evidence is similar to that observed in general medicine (Ellis *et al*, 1995; Geddes *et al*, 1996*b*; Summers & Kehoe, 1996). However, it could be argued that these data only apply to 'medical' type interventions and that most psychiatry lies in 'grey zones' of clinical practice which lie outside the scope of EBM (Tanenbaum, 1993; Naylor, 1995). Indeed, these grey zones may be considered to be the essence of psychiatry since they include the nature of the doctor-patient relationship, therapist qualities, and even the effectiveness of compulsory treatment. While we are not suggesting that everything in psychiatry can currently be based soundly on evidence which we can find if only we know where to look (see Greenhalgh, 1996), we would argue, in line with Sullivan & MacNaughton (1996), that the interview and other key elements of psychiatry *are* amenable to an evidence-based approach. For example, one could carry out an RCT to establish whether scrupulously balanced prognostic statements or unfailingly optimistic (but not untrue) ones lead to better clinical outcomes.

EBM will be used an an excuse to cut services and limit clinical freedom

There is an understandable anxiety that purchasers, under the guise of evidence-based purchasing, will attempt to disinvest in

JOHN R. GEDDES, MRCPsych, PAUL J. HARRISON, MRCPsych University Department of Psychiatry, Warneford Hospital, Oxford

Correspondence: Dr J. Geddes, University Department of Psychiatry, Warneford Hospital, Oxford OX37JX. Fax: 01865793101; e-mail:john.geddes@psychiatry.oxford.ac.uk

(First received 18 November 1996, final revision 20 November 1996, accepted 1 May 1997)

established services for which there is no high quality (RCT) evidence of effectiveness (Grahame-Smith, 1995). It is easy to accept the aim of EBM to increase awareness and implementation of genuine treatment advances. Equally, hallowed procedures must be abandoned with the same enthusiasm if evidence of their relative ineffectiveness or actual harm emerges (Eddy, 1994) as, for example, in the case of high-dose antipsychotics (Hirsch & Barnes, 1994).

As health care rationing becomes explicit, battles are likely to be fought over 'old' treatments which were introduced before the current requirements for evidence of effectiveness (Eddy, 1993). It is over these interventions that disagreements are most likely to occur between providers and purchasers, since the two groups have different priorities (Eddy, 1992): clinicians should stress that absence of RCT evidence of effectiveness is not the same as evidence of ineffectiveness, and should take the view that the burden of proof lies with those who wish to disinvest in an established treatment. Informed debate about service provision will be facilitated by the use of EBM. Moreover, if clinicians are able to argue effectively for the introduction of evidence-based interventions, EBM may lead to increased costs rather than to savings (Sackett *et al*, 1996*a*).

Finally, there has been concern about the loss of clinical freedom because of EBM. However, it is only when the term is misappropriated to mean the right to carry on practising demonstrably inferior medicine that EBM—rightly—restricts clinical freedom.

Conclusion

Psychiatrists must keep abreast of therapeutic advances, cope with rapidly changing mental health policies, and face increasing public expectations and demands from purchasers. We would argue that the best response to these pressures is to use EBM, since EBM optimises our clinical decisions and allows us to justify them. We would also suggest that it makes practice easier and more efficient. As with the implications of recent scientific advances, the changes in practice which EBM requires are challenging. However, the EBM approach to clinical decision-making is like a scalpel compared with the meat-axe of top–down policy-making (Eddy, 1990). Psychiatrists will be better off adopting EBM and applying it judiciously, rather than waiting for it to be foisted upon us.

References

Anonymous (1995) Care-management: a disastrous mistake. *Lancet*, 345, 399–401.

Bennett, K. J., Sackett, D. L., Haynes, R. B., *et al* (1937) A controlled trial of teaching critical appraisal of the clinical literature to medical students. *Journal of the American Medical Association*, 257: 2451–2454.

Beral, V., Hermon, C., Reeves, G., *et al*, (1995) Sudden fall in breast cancer death rates in England and Wales. *Lancet*, 345: 1642–1643.

Cochrane Collaboration (1996) *The Cochrane Library*. Oxford: Update Software.

Cook, R. J. & Sackett, D. L. (1995) The number needed to treat: a clinically useful measure of treatment effect. *British Medical Journal*, 310: 452–454.

Davis, D. A., Thomson, M. A., Oxman, A. D., *et al* (1995) Changing physician performance. A systematic review of the effect of continuing medical education strategies. *Journal of the American Medical Association*, 274: 700–705.

Eccles, M., Clapp, Z., Grimshaw, J., *et al* (1996) North of England evidence based guidelines development project: methods of guideline development. *British Medical Journal*, 312: 760–762.

Eddy, D. M. (1990) Clinical decision making: from theory to practice. What do we do about costs? *Journal of the American Medical Association*, 264: 1161,1165,1169–1170.

—— (1992) Clinical decision making: from theory to practice. Cost-effectiveness analysis. Will it be accepted? *Journal of the American Medical Association*, 268: 132–136.

—— (1993) Clinical decision making: from theory to practice. Three battles to watch in the 1990s. *Journal of the American Medical Association*, 270: 520–526.

—— (1994) Clinical decision making: from theory to practice. Prinicples for making difficult decision in difficult times. *Journal of American Medical Association*, 271: 1792–1798.

Egger, M. & Smith, G. D. (1995) Misleading meta-analysis. *British Medical Journal*, 310: 752–754.

Ellis,J., Mulligan, I., Rowe, J., *et al* (1995) Inpatient general medicine is evidence based. *Lancet*, 346: 407–410.

Evidence-Based Medicine Working Group (1992) Evidence-based medicine. A new approach to teaching the practice of medicine. *Journal of the American Medical Association*, 268: 2420–2425.

Eysenck, H. J.(1994) Meta-analysis and its problems. *British Medical Journal*, 309: 789–792.

Fahey, T., Griffiths, S. & Peters, T. J. (1995) Evidence based purchasing: understanding results of clinical trials and systematic reviews. *British Medical Journal*, 311: 1056–1059.

Geddes, J. R., Christofi, G. & Sackett, D. L. (1999*a*) Commentary on "first-rank symptoms", *British Journal of Psychiatry*, 169: 544–545.

Geddes, J. R., Game, D., Jenkins, N. E., *et al* (1996*b*) What proportion of primary psychiatric interventions are based on randomised evidence? *Quality in Health Care*, 5: 215–217.

Gilbody, S. (1996) Evidence-based medicine: an improved format for journal clubs. *Psychiatric Bulletin*, 20: 673–675.

Glasziou, P. P. & Irwig, L. M. (1995) An evidence based approach to individualising treatment. *British Medical Journal*, 311: 1356–1359.

Grahame Smith, D. (1995) Evidence based medicine: Socratic dissent. *British Medical Journal*, 310: 1126–1127.

Greenhalgh, T. (1996) Is my practice evidence-based? *British Medical Journal*, 313: 957–958.

Hermann, R. C., Dorwart, R. A., Hoover, C.W., *et al* (1995) Variation in ECT use in the United States. *American Journal of Psychiatry*, 152: 869–875.

Hirsch, S. R. & Barnes, T. R. (1994) Clinical use of high-dose neuroleptics. *British Journal of Psychiatry*, 164: 94–96.

Huston, P. & Moher, D. (1996) Redundancy, disaggregation, and the integrity of medical research. *Lancet*, 347: 1024–1026.

Jackson, R. T. & Sackett, D. L. (1996) Guidelines for managing raised blood pressure. *British Medical Journal*, 313: 64–65.

Kerr, M., Blizard, R. & Mann, A. (1995) General practitioners and psychiatrists: comparison of attitudes to depression using the depression attitude questionnaire. *British Journal of General Practice*, 45: 89–92.

Laupacis, A., Sackett, D. L. & Roberts, R. S. (1988) An assessment of clinically useful measures of the consequences of treatment. *New England Journal of Medicine*, 318: 1728–1733.

Lewis, A. (1958) Between guesswork and uncertainty in psychiatry. *Lancet, i*: 171–175.

Marshall, M., Gray, A., Lockwood, A., et al (1996) Case management for people with severe mental disorders. In *Schizophrenia Module* (eds C. Adams, J. Anderson & J. De Jesus Mari). Oxford: Update Software.

Meise, U., Kurz, M. & Fleischhacker, W. W. (1994) Antipsychotic maintenance treatment of schizophrenia patients: is there a consensus? *Schizophrenia Bulletin*, 20: 215–225.

Mulrow, C. D. (1987) The medical review article: state of the science. *Annals of Internal Medicine*, 106: 485–488.

—— (1994) Rationale for systematic reviews. *British Medical Journal*, 309: 597–599.

Naylor, C. D. (1995) Grey zones of clinical practice: some limits to evidence-based medicine. *Lancet*, 345: 840–842.

NHS Centre for Reviews and Dissemination (1994) Implementing Clinical Practice Guidelines. *Effective Health Care*, 8.

Oxman, A. D. & Guyatt, G. H. (1988) Guidelines for reading literature reviews. *Canadian Medical Association Journal*, 138: 679–703.

Pippard, J. (1992) Audit of electroconvuslive treatment in two national health service regions. *British Journal of Psychiatry*, 160: 621–637.

Priest, R. G., Vize, C., Roberts, A., *et al* (1996) Lay people's attitudes to treatment of depression: results of opinion poll for Defeat Depression Campaign just before its launch. *British Medical Journal*, 313: 858–859.

Sackett, D. L., Rosenberg, W. M., Gray, J. A., *et al* (1996*a*) Evidence based medicine: what it is and what it isn't. *British Medical Journal*, 312: 71–72.

—— Deeks, J. J. & Altman, D. G. (1996*b*) Down with odds ratios! (EBM note). *Evidence-Based Medicine*, 1: 164–166.

——, Richardson, S., Rosenberg, W., *et al* (1997) *Evidence-Based Medicine*: *How to Practise and Teach EBM*. London: Churchill-Livingstone.

Shahar, E. (1995) Evidence-based medicine. *Lancet*, 346: 1772.

Sheldon, T. A. (1996) Research intelligence for policy and practice: the role of the National Health Service Centre for Reviews and Dissemination (EBM note). *Evidence-Based Medicine*, 1: 167–168.

Shin, J. H., Haynes,R.B. & Johnston, M. E. (1993) Effect of problem-based, self-directed undergraduate education on life-long learning. *Canadian Medical Association Journal*, 148: 969–976.

Sim, I. & Hlatky, M. A. (1996) Growing pains of metaanalyses. *British Medical Journal*, 313: 702–703.

Simon, G., Wagner, E. & Wonkorff, M. (1995) Cost-effectiveness comparisions using "real world" randomized trials: the case of new antidepressant drugs. *Journal of Clinical Epidemiology*, 48: 363–373.

Smith, R. (1991) Where is the wisdom . . . ? *British Medical Journal*, 303, 798–799.

—— (1996) What clinical information do doctors need? *British Medical Journal*, 313: 1062–1068.

Sullivan, F. M. & MacNaughton, R. J. (1996) Evidence in consultations: interpreted and individualised. *Lancet*, 348: 941–943.

Summers, A. & Kehoe, R. F. (1996) Is psychiatric treatment evidence-based? *Lancet*, 347: 409–410.

Tanenbaum, S. J. (1993) What physicians know. *New England Journal of Medicine*, 329: 1268–1271.

Wells, K. B., Katon, W., Rogers, B., *et al* (1994) Use of minor tranquilizers and antidepressant medications by depressed outpatients: results from the medical outcomes study. *American Journal of Psychiatry*, 151: 694–700.

Reading 16.2

EXERCISE 2

From: Hume, D. ([1748] 1975). *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*. Oxford: Oxford University Press, section IV (Extract pp. 37–8).

When a man says, *I have found, in all past instances, such sensible qualities conjoined with such secret powers*: And when he says, *Similar sensible qualities will always be conjoined with similar secret powers*, he is not guilty of a tautology, nor are these propositions in any respect the same. You say that the one proposition is an inference from the other. But you must confess that the inference is not intuitive; neither is it demonstrative: Of what nature is it, then? To say it is experimental, is begging the question. For all inferences from experience suppose, as their foundation, that the future will resemble the past, and that similar powers will be conjoined with similar sensible qualities. If there be any suspicion that the course of nature may change, and that the past may be no rule for the future, all experience becomes useless, and can give rise to no inference or conclusion. It is impossible, therefore, that any arguments from experience can prove this resemblance of the past to the future; since all these arguments are founded on the supposition of that resemblance. Let the course of things be allowed hitherto ever so regular; that alone, without some new argument or inference, proves not that, for the future, it will continue so. In vain do you pretend to have learned the nature of bodies from your past experience. Their secret nature, and consequently all their effects and influence, may change, without any change in their sensible qualities. This happens sometimes, and with regard to some objects: Why may it not happen always, and with regard to all objects? What logic, what process of argument secures you against this supposition? My practice, you say, refutes my doubts. But you mistake the purport of my question. As an agent, I am quite satisfied in the point; but as a philosopher, who has some share of curiosity, I will not say scepticism, I want to learn the foundation of this inference. No reading, no enquiry has yet been able to remove my difficulty, or give me satisfaction in a matter of such importance. Can I do better than propose the difficulty to the public, even though, perhaps, I have small hopes of obtaining a solution? We shall at least, by this means, be sensible of our ignorance, if we do not augment our knowledge.

Reading 16.3

EXERCISE 4

From: Popper, K. (1972). Conjectural knowledge: my solution to the problem of induction. In *Objective Knowledge*. Oxford: Oxford University Press, pp. 1–31 (Extract pp. 7–9).

5. The Logical Problem of Induction: Restatement and Solution

In accordance with what has just been said (point (2) of the preceding section 4), I have to restate Hume's H_L in an objective or logical mode of speech.

To this end I replace Hume's 'instances of which we have experience' by 'test statements'—that is, singular statements describing observable events ('observation statements', or 'basic statements'); and 'instances of which we have no experience' by 'explanatory universal theories'.

I formulated Hume's logical problem of induction as follows:

L_1 Can the claim that an explanatory universal theory is true be justified by 'empirical reasons'; that is, by assuming the truth of certain test statements or observation statements (which, it may be said, are 'based on experience')?

My answer to the problem is the same as Hume's: No, we cannot; no number of true test statements would justify the claim that an explanatory universal theory is true.[1]

But there is a second logical problem, L_2, which is a generalization of L_1. It is obtained from L_1 merely by replacing the words 'is true' by the words 'is true or that it is false':

L_2 Can the claim that an explanatory universal theory is true or that it is false be justified by 'empirical reasons': that is, can the assumption of the truth of test statements justify either the claim that a universal theory is true or the claim that it is false?

To this problem, my answer is positive: Yes, *the assumption of the truth of test statements sometimes allows us to justify the claim that an explanatory universal theory is false*.

This reply becomes very important if we reflect on the problem situation in which the problem of induction arises. I have in mind a situation in which we are faced with *several explanatory theories* which compete *qua* solutions of some problem of explanation—for example a scientific problem; and also with the fact that we have to, or at least wish to, choose between them. As we have seen, Russell says that without solving the problem of induction, we could not *decide between* a (good) scientific theory and a (bad) obsession of a madman. Hume too had competing theories in mind. 'Suppose [he writes] a person . . . advances propositions, to which I do not assent, . . . that silver is more fusible than lead, or mercury heavier than gold. . . .'[2]

This problem situation—that of choosing between several theories—suggests a third reformulation of the problem of induction:

L_3 Can a *preference*, with respect to truth of falsity, for some competing universal theories over others ever be justified by such ' empirical reasons'?

In the light of my answer to L_2 the answer to L_3 becomes obvious: Yes; sometimes it can, if we are lucky. For it may happen that our test statements may refute some—but not all—of the competing theories; and since we are searching for a true theory, we shall prefer those whose falsity has not been established.

6. Comments on My Solution of the Logical Problem

(1) According to my reformulations, the central issue of the logical problem of induction is the validity (truth or falsity) of universal laws *relative to some 'given' test statements.* I do not raise the question, 'How do we decide the truth or falsity of test statements?', that is, of singular descriptions of observable events. The latter question should not, I suggest, be regarded as part of the problem of induction, since Hume's question was whether we are justified in reasoning from experienced to unexperienced 'instances'.[3] Neither Hume nor any other writer on the subject before me has to my knowledge moved on from here to the *further questions*: Can we take the 'experienced instances' for granted? And are they really prior to the theories? Although these further questions are some of those problems to which I was led by my solution of the problem of induction, they go beyond the original problem. (This is clear if we consider the kind of thing for which philosophers have been looking when trying to solve the problem of induction: if a 'principle of induction', permitting us *to derive universal laws from singular statements*, could be found, and its claim to truth defended, then the problem of induction would be regarded as solved.)

(2) L_1 is an attempt to translate Hume's problem into an objective mode of speech. The only difference is that Hume speaks of future (singular) *instances* of which we have no experience—that is, of expectations—while L_1 speaks of universal laws or theories. I have at least three reasons for this change. First, from a logical point of view, 'instances' are relative to some universal law (or at least to a statement function which

[1] An explanatory theory goes essentially beyond even an infinity of universal test statements; even a law of low universality does so.

[2] Hume, Treatise, p. 95.

[3] Op. cit., p. 91.

could be universalized). Secondly, our usual method of reasoning from 'instances' to other 'instances' is with the help of universal theories. Thus we are led from Hume's problem to the *problem of the validity of universal theories* (their truth or falsehood). Thirdly, I wish, like Russell, to connect the problem of induction with *the universal laws or theories of science.*

(3) My negative answer to L_1 should be interpreted as meaning that we must regard *all laws or theories as hypothetical or conjectural*; that is, as guesses.

Reading 16.4

EXERCISE 5

From: Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In *Criticism and the Growth of Knowledge* (ed. I. Lakatos and A. Musgrave). Cambridge: Cambridge University Press, pp. 91–195 (Extract pp. 100–101).

The story is about an imaginary case of planetary misbehaviour. A physicist of the pre-Einsteinian era takes Newton's mechanics and his law of gravitation (N), the accepted initial conditions, I, and calculates, with their help, the path of a newly discovered small planet, p. But the planet deviates from the calculated path. Does our Newtonian physicist consider that the deviation was forbidden by Newton's theory and therefore that, once established, it refutes the theory N? No. He suggests that there must be a hitherto unknown planet p' which perturbs the path of p. He calculates the mass, orbit, etc., of this hypothetical planet and then asks an experimental astronomer to test his hypothesis. The planet p' is so small that even the biggest available telescopes cannot possibly observe it: the experimental astronomer applies for a research grant to build yet a bigger one. In three years' time the new telescope is ready. Were the unknown planet p' to be discovered, it would be hailed as a new victory of Newtonian science. But it is not. Does our scientist abandon Newton's theory and his idea of the perturbing planet? No. He suggests that a cloud of cosmic dust hides the planet from us. He calculates the location and properties of this cloud and asks for a research grant to send up a satellite to test his calculations. Were the satellite's instruments (possibly new ones, based on a little-tested theory) to record the existence of the conjectural cloud, the result would be hailed as an outstanding victory for Newtonian science. But the cloud is not found. Does our scientist abandon Newton's theory, together with the idea of the perturbing planet and the idea of the cloud which hides it? No. He suggests that there is some magnetic field in that region of the universe which disturbed the instruments of the satellite. A new satellite is sent up. Were the magnetic field to be found, Newtonians would celebrate a sensational victory. But it is not. Is this regarded as a refutation of Newtonian science? No. Either yet another ingenious auxiliary hypothesis is proposed or . . . the whole story is buried in the dusty volumes of periodicals and the story never mentioned again.

Reading 16.5 EXERCISE 6

From: Kuhn, T. S. (1970). Logic or discovery or psychology of research? In *Criticism and the Growth of Knowledge* (ed. I. Lakatos and A. Musgrave). Cambridge: Cambridge University Press, pp. 1–23 (Extract pp. 4–6).

Among the most fundamental issues on which Sir Karl and I agree is our insistence that an analysis of the development of scientific knowledge must take account of the way science has actually been practiced. That being so, a few of his recurrent generalizations startle me. One of these provides the opening sentences of the first chapter of the *Logic of Scientific Discovery*: 'A scientist', writes Sir Karl, 'whether theorist or experimenter, puts forward statements, or systems of statements, and tests them step by step. In the field of the empirical sciences, more particularly, he constructs hypotheses, or systems of theories, and tests them against experience by observation and experiment.'[1] The statement is virtually a cliché, yet in application it presents three problems. It is ambiguous in its failure to specify which of two sorts of 'statements' or 'theories' are being tested. That ambiguity can, it is true, be eliminated by reference to other passages in Sir Karl's writings, but the generalization that results is historically mistaken. Furthermore, the mistake proves important, for the unambiguous form of the description misses just that characteristic of scientific practice which most nearly distinguishes the sciences from other creative pursuits.

There is one sort of 'statement' or 'hypothesis' that scientists do repeatedly subject to systematic test. I have in mind statements of an individual' s best guesses about the proper way to connect his own research problem with the corpus of accepted scientific knowledge. He may, for example, conjecture that a given chemical unknown contains the salt of a rare earth, that the obesity of his experimental rats is due to a specified component in their diet, or that a newly discovered spectral pattern is to be understood as an effect of nuclear spin. In each case, the next steps in his research are intended to try out or test the conjecture or hypothesis. If it passes enough or stringent enough tests, the scientist has made a discovery or has at least resolved the puzzle he had been set. If not, he must either abandon the puzzle entirely or attempt to solve it with the aid of some other hypothesis. Many research problems, though by no means all, take this form. Tests of this sort are a standard component of what I have elsewhere labelled 'normal science' or 'normal research', an enterprise which accounts for the overwhelming majority of the work done in basic science. In no usual sense, however, are such tests directed to current theory. On the contrary, when engaged with a normal research problem, the scientist must *premise* current theory as the rules of his game. His object is to solve a puzzle, preferably one at which others have failed, and current theory is required to define that puzzle and to guarantee that, given sufficient brilliance, it can be solved.[2] Of course the practitioner of such an enterprise must often test the conjectural puzzle solution that his ingenuity suggests. But only his personal conjecture is tested. If it fails the test, only his own ability not the corpus of current science is impugned. In short, though tests occur frequently in normal science, these tests are of a peculiar sort, for in the final analysis it is the individual scientist rather than current theory which is tested.

This is not, however, the sort of test Sir Karl has in mind. He is above all concerned with the procedures through which science grows, and he is convinced that 'growth' occurs not primarily by accretion but by the revolutionary overthrow of an accepted theory and its replacement by a better one.[3] (The subsumption under 'growth' of 'repeated overthrow' is itself a linguistic oddity whose *raison d'être* may become more visible as we proceed.) Taking this view, the tests which Sir Karl emphasizes are those which were performed to explore the limitations of accepted theory or to subject a current theory to maximum strain. Among his favourite examples, all of them startling and destructive in their outcome, are Lavoisier's experiments on calcination, the eclipse expedition of 1919, and the recent experiments on party conservation.[4] All, of course, are classic tests, but in using them to characterize scientific activity Sir Karl misses something terribly important about them. Episodes like these are very rare in the development of science. When they occur, they are generally called forth either by a prior crisis in the relevant field (Lavoisier's experiments or Lee and Yang's[5] or by the existence of a theory which competes with the existing canons of research (Einstein's general relativity). These are, however, aspects of or occasions for what I have elsewhere called 'extraordinary research', an enterprise in which scientists do displây very many of the characteristics Sir Karl emphasizes, but one which, at least in the past, has arisen only intermittently and under quite special circumstances in any scientific speciality.[6]

I suggest then that Sir Karl has characterized the entire scientific enterprise in terms that apply only to its occasional revolutionary parts. His emphasis is natural and common: the exploits of a Copernicus or Einstein make better reading than those of a Brahe or Lorentz; Sir Karl would not be the first if he mistook what

[1] Popper [1959], p. 27.

[2] For an extended discussion of normal science, the activity which practitioners are trained to carry on, see my [1962], pp. 23–42, and 135–42. It is important to notice that when I describe the scientist as a puzzle solver and Sir Karl describes him as a problem solver (e.g. in his [1963], pp. 67, 222), the similarity of our terms disguises a fundamental divergence. Sir Karl writes (the italics are his), 'Admittedly, our expectations, and thus our theories, may precede, historically, even our problems. *Yet science starts only with problems*. Problems crop up especially when we are disappointed in our expectations, or when our theories involve us in difficulties, in contradictions'. I use the term 'puzzle' in order to emphasize that the difficulties which *ordinarily* confront even the very best scientists are, like crossword puzzles or chess puzzles, challenges only to his ingenuity. *He* is in difficulty, not current theory. My point is almost the converse of Sir Karl's.

[3] Cf. Popper [1963], pp. 129, 215 and 221, for particularly forceful statements of this position.

[4] For example, Popper [1963], p. 220.

[5] For the work on calcination see, Guerlac [1961]. For the background of the parity experiments see, Hafner and Presswood [1965].

[6] The point is argued at length in my [1962], pp. 52–97.

I call normal science for an intrinsically uninteresting enterprise. Nevertheless, neither science nor the development of knowledge is likely to be understood if research is viewed exclusively through the revolutions it occasionally produces. For example, though testing of basic commitments occurs only in extraordinary science, it is normal science that discloses both the points to test and the manner of testing. Or again, it is for the normal, not the extraordinary practice of science that professionals are trained; if they are nevertheless eminently successful in displaying and replacing the theories on which normal practice depends, that is an oddity which must be explained. Finally, and this is for now my main point, a careful look at the scientific enterprise suggests that it is normal science, in which Sir Karl's sort of testing does not occur, rather than extraordinary science which most nearly distinguishes science from other enterprises. If a demarcation criterion exists (we must not, I think, seek a sharp or decisive one), it may lie just in that part of science which Sir Karl ignores.

Reading 16.6

EXERCISE 7

From: Bloor, D. (1976). *Knowledge and Social Imagery*. London: Routledge, (Extract pp. 1–3).

Can the sociology of knowledge investigate and explain the very content and nature of scientific knowledge? Many sociologists believe that it cannot. They say that knowledge as such, as distinct from the circumstances surrounding its production, is beyond their grasp. They voluntarily limit the scope of their own enquiries. I shall argue that this is a betrayal of their disciplinary standpoint. All knowledge, whether it be in the empirical sciences or even in mathematics, should be treated, through and through, as material for investigation. Such limitations as do exist for the sociologist consist in handing over material to allied sciences like psychology or in depending on the researches of specialists in other disciplines. There are no limitations which lie in the absolute or transcendent character of scientific knowledge itself, or in the special nature of rationality, validity, truth or objectivity.

It might be expected that the natural tendency of a discipline such as the sociology of knowledge would be to expand and generalise itself: moving from studies of primitive cosmologies to that of our own culture. This is precisely the step that sociologists have been reluctant to take. Again, the sociology of knowledge might well have pressed more strongly into the area currently occupied by philosophers, who have been allowed to take upon themselves the task of defining the nature of knowledge. In fact sociologists have been only too eager to limit their concern with science to its institutional framework and external factors relating to its rate of growth or direction. This leaves untouched the nature of the knowledge thus created (cf. Ben-David (1971), DeGré (1967), Merton (1964) and Stark (1958)).

What is the cause for this hesitation and pessimism? Is it the enormous intellectual and practical difficulties which would attend such a programme? Certainly these must not be underestimated. A measure of their extent can be gained from the effort that has been expended on the more limited aims. But these are not the reasons that are in fact advanced. Is the sociologist at a loss for theories and methods with which to handle scientific knowledge? Surely not. His own discipline provides him with exemplary studies of the knowledge of other cultures which could be used as models and sources of inspiration. Durkheim's classic study 'The Elementary Forms of the Religious Life' shows how a sociologist can penetrate to the very depths of a form of knowledge. What is more Durkheim dropped a number of hints as to how his findings might relate to the study of scientific knowledge. The hints have fallen on deaf ears.

The cause of the hesitation to bring science within the scope of a thorough-going sociological scrutiny is lack of nerve and will. It is believed to be a foredoomed enterprise. Of course, the failure of nerve has deeper roots than this purely psychological characterisation suggests, and these will be investigated later. Whatever the cause of the malady, its symptoms take the form of a priori and philosophical argumentation. By these means sociologists express their conviction that science is a special case, and that contradictions and absurdities would befall them if they ignored this fact. Naturally philosophers are only too eager to encourage this act of self-abnegation (e.g. Lakatos (1971), Popper (1966)).

It will be the purpose of this book to combat these arguments and inhibitions. For this reason the discussions which follow will sometimes, though not always, have to be methodological rather than substantive. But hopefully they will be positive in their effect. Their aim is to put weapons in the hands of those engaged in constructive work to help them attack critics, doubters and sceptics.

I shall first spell out what I call the strong programme in the sociology of knowledge. This will provide the framework within which detailed objections will then be considered. Since a priori arguments are always embedded in background assumptions and attitudes it will be necessary to bring these to the surface for examination as well. This will be the second major topic and it is here that substantial sociological hypotheses about our conception of science will begin to emerge. The third major topic will concern what is perhaps the most difficult of all the obstacles to the sociology of knowledge, namely mathematics and logic. It will transpire that the problems of principle involved are not, in fact, unduly technical. I shall indicate how these subjects can be studied sociologically.

The strong programme

The sociologist is concerned with knowledge, including scientific knowledge, purely as a natural phenomenon. His definition of knowledge will therefore be rather different from that of either the layman or the philosopher. Instead of defining it as true belief, knowledge for the sociologist is whatever men take to be knowledge. It consists of those beliefs which men confidently hold to and live by. In particular the sociologist will be concerned with beliefs which are taken for granted or institutionalised, or invested with authority by groups of men. Of course knowledge must be distinguished from mere belief. This can be done by reserving the word 'knowledge' for what is collectively endorsed, leaving the individual and idiosyncratic to count as mere belief.

Men's ideas about the workings of the world have varied greatly. This has been true within science just as much as in other areas of culture. Such variation forms the starting point for the sociology of knowledge and constitutes its main problem. What are the causes of this variation, and how and why does it change? The sociology of knowledge focuses on the distribution of belief and the various factors which influence it. For example: how is knowledge transmitted; how stable is it; what processes go into its creation and maintenance; how is it organised and categorised into different disciplines or spheres?

For the sociologist these topics call for investigation and explanation and he will try to characterise knowledge in a way which accords with this perspective. His ideas therefore will be in the same causal idiom as any other scientist. His concern will be to locate the regularities and general principles or processes which appear to be at work within the field of his data. His aim will be to build theories to explain these regularities. If these theories are to satisfy the requirement of maximum generality they will have to apply to both true and false beliefs, and as far as possible the same type of explanation will have to apply in both cases. The aim of physiology is to explain the organisam in health and disease; the aim of mechanics is to understand machines which work and machines which fail; bridges which stand as well as those which fall. Similarly the sociologist seeks theories which explain the beliefs which are in fact found, regardless of how the investigator evaluates them.

Reading 16.7

EXERCISE 8

From: Mellor, D.H. (1991). The warrant of induction. In *Matters of Metaphysics*. London: Routledge. (Extract sections 1 and 6).

Extract 1

1 Introduction

This lecture will last less than twenty-four hours. I know that, and so do you. And you knew it before I said so. How? Because you knew that lectures don't last twenty-four hours. How do you know that? You haven't heard this one, and 'for all you know' (as the saying is) I could go on all night. But you know I won't. And the 'all you know' which tells you that, without entailing it, is the fact that none, or almost none, of the many lectures, on all subjects, which you've heard or heard of, have lasted that long. If many of them had, you wouldn't have known that this one won't; but as it is, you do know that.

That's a piece of induction. We believe that something (a lecture) has an observable but as yet unobserved property (a short duration), and this belief is warranted by the fact that many otherwise diverse things of the same kind (lectures) have all—or nearly all—been observed to have that property. The problem, set for us by David Hume in 1739, is to say why such observations warrant such a belief.

The greatest of all Cambridge philosophers, Frank Ramsey, said in 1926 that 'we are all convinced by inductive arguments, and our conviction is reasonable because the world is so constituted that inductive arguments lead on the whole to true opinions'. As usual, he was right, but too brisk for most philosophers. Most are still unpersuaded, despite the subsequent work of Professor Braithwaite and others. Tonight I will try to say, less briskly, but still within twenty-four hours, why I think Ramsey was right.

First I must say what it is to warrant a belief. Or rather, what it is about a belief that is to be warranted. What's to be warranted is its truth, that is the truth of what's believed: for example, that this lecture will last less than twenty-four hours. That's why observations which warrant a belief can't also warrant a contradictory one: contradictory beliefs can't both be true. Yet our inductive warrant doesn't entail that the belief it warrants is true. The belief that this lecture will stop tonight would be warranted now even if I went on to make that belief false; and the belief I would then make true would still not be warranted now. But how can a warrant be valid and yet fail? That's part of the problem of induction.

The problem isn't just that your inductive evidence fails to entail the belief it warrants. It's worse than that. For that belief is a prediction, about when this lecture will end. And its success as a prediction will be settled in due course by an observation: for example, by your consulting a watch when the lecture ends. Some such observation will settle that prediction regardless of its present inductive warrant. If I keep you here till dawn, your present inductive evidence that I won't will count for nothing against the evidence of your senses that I have done. And if I stop sooner, your senses will need no help from induction to tell you so. You know that your inductive warrant for what you now believe is negligible compared with the warrant that some future observation will give it—or deny it.

Yet we trust such predictive beliefs incessantly, often with our lives. Take the belief that this building will stay up while I speak two more sentences. We have, I hope, strong inductive warrant for that belief, and thus for staying here at least that long. Yet the belief is still only a prediction (comma), because we know that its present warrant will be quite superseded by that of our senses . . . now. But how can a warrant that is now so weak ever have been strong enough to trust our lives to? And as in this case, so in general: how can a prediction ever be better than a guess? That's another way of putting the problem of induction.

Extract 2

6 Induction

Let me call events which last less than twenty-four hours 'terse'. This event, as you see, is a lecture. It's also terse, but that you can't yet see. You predict it, by inferring it: this is a lecture, so it's terse. And this inference displays a general disposition or habit: you would have drawn it of any lecture, not just this one. And this inferential disposition, like your mass (your inertial disposition), embodies a causal link: just as your mass makes forces cause you to accelerate, so your inferential disposition makes your coming to believe that an event is a lecture cause you to predict that it's terse.

We've already seen how causal links can enable observations to warrant beliefs: namely, when the fact that makes a belief true is what causes you to get that belief. The role of causation in an inference is different. What causes you to believe the conclusion of an inference isn't the fact that makes it true, but the fact that you've come to believe the premise of the inference. How can this warrant the conclusion?

Suppose, to simplify the discussion (it's not essential), that your inferential habits are deterministic, like your mass: this habit would always make you infer that a lecture was terse, never that it wasn't. And suppose that every lecture has some chance of being terse. Then whenever your premise ('this is a lecture') is true, your conclusion ('this is terse') has some chance of being true. And if this chance is high enough, your prediction is warranted.

I shall call an inferential habit warranted (or goad) if, whenever the premise is true, the conclusion has a high enough chance of being true. For then there's a very high chance that all or nearly all the conclusions of many such inferences will be true when their premises are. A particular inference may still of course fail to warrant its conclusion, because its premise may be false. But that's no fault of the inference. All an inference can do is give a conclusion as high a chance of being warranted as its premise has of being true. That's good enough to warrant calling it warranted, or good.

We have of course many habits of inference. Some perhaps we're born with; most we acquire. And, as Hume remarked, we acquire most of them by induction. Mostly, the more we see that otherwise diverse things with one property F (being a lecture) all or nearly all have another property G (being terse), the more we tend to predict that other things which are F are also G.

This isn't a universal tendency: it doesn't apply to all observable properties. Take ageing. The more years we adults do survive, the more we tend to predict that we won't survive another one. We recognise many such 'counter-inductive' phenomena: metal fatigue and caterpillars turning into butterflies are two more obvious examples. But we recognise them inductively. Our mortality statistics show that the observed chances of adults surviving another year have so far always or nearly always decreased with age. That's why we tend to predict that ours will. And similarly in the other cases. Our counter-inductive tendencies always have an inductive basis.

The question is: what warrants the inductive basis? We certainly think it's warranted. Counter-inductivists aren't just odd: they're mad. Imagine one. He won't eat bread: he thinks it would poison him, because it's never poisoned anyone before. He would eat cyanide, which he also expects to freeze in the oven and bake in the fridge; but not by swallowing it. He won't use any language people have so far understood, or breathe air, or drink water. And throughout his (brief) life he consistently defends his wholesale counter-inductivism by predicting that as it's almost never worked yet, it will now.

And so it could. He could be right. He could outlive us all. But he won't; and we know he won't. How? What makes induction a better basic tendency than counter-induction?

Take any pair of basic properties which we've seen to be correlated, like being a frog and being green. Suppose all or nearly all the many frogs so far seen have been green. The more, and the more diverse, those frogs and their surroundings, the more we tend to predict that other frogs will be green. Exactly how many, and how diverse, doesn't matter: what matters is what warrants this general inductive tendency.

The tendency increases both with the number of frogs seen, and with their diversity. Consider first the way it increases with the number. So suppose the frogs (and their surroundings) are all of a kind: tree frogs, for example. The more tree frogs are all seen to be green, the more we tend to predict that other tree frogs are green—and the more counter-inductivists tend to predict that they're not. What makes ours the better tendency?

That depends on what chance a tree frog actually has of being green. It may be a law of nature that all tree frogs are green: that is, the chance of any tree frog being green may be 1. Suppose it is. Then all observed tree frogs will be green, so induction will always make us predict that other tree frogs are green. And this inference couldn't be better: whenever its premise ('this is a tree frog') is true, the chance of its conclusion ('this is green') being true is 1. Whereas counter-inductivists, seeing only green tree frogs, will always predict that other tree frogs aren't green: an inference which couldn't be worse.

So much the worse for counter-induction. And it's no better off if the law is that no tree frogs are green. For then no observed tree frogs will be green, induction will always yield the good habit of inferring that others aren't either, and counter-induction the bad habit of inferring that they are. And as for frogs and colour, so for all basic observable properties. Whenever they're linked by a deterministic law, induction will always yield good habits of inference, and counter-induction will always yield bad ones. In all such cases our inductive tendency is warranted, and the counter-inductive one isn't.

So far so good—provided these warrants needn't be self-intimating. I say the law that all tree frogs are green warrants my inferring that something is green from the fact that it's a tree frog. But suppose I must know that I have this warrant. Then I must know this law. So I must believe it, and this belief must be warranted. But the law entails the very inference which it's meant to warrant: tree frogs can't all be green unless this one is. So unless my inference is warranted already, my belief in the law won't be warranted. Thus to claim that the law is what warrants this particular application of it simply begs the question of whether it's warranted at all.

This is the stock objection to contingent solutions to the problem of induction: they beg the question. And so they would if belief-warrants had to be self-intimating. But as we've seen, they don't. The law that all tree frogs are green can warrant the habit of inference which induction will then give me, just because I needn't know that it does. I can know by induction that a frog is green without knowing that law, just as I can know that it's green by looking at it without knowing I'm not colour-blind. I may know the law, just as I may know that I'm not colour-blind; but I needn't. So my saying that the law is what warrants this induction doesn't beg the question.

Deterministic laws warrant induction. And so do statistical laws, and hence the chances that entail them. For suppose it's a law that all tree frogs have a certain chance of being green. The greater this chance, the better the inference that a tree frog is green. But also the greater the chance that many tree frogs will all or nearly all be green. In particular, the greater the chance that all or nearly all observed tree frogs will be green—and hence that we inductivists will infer that others are too. In short, the better the inference, the more likely we inductivists are to make it. And the less likely counter-inductivists are (since the better it is, the less chance there is of all or nearly all observed tree frogs not being green). So far still so good.

Furthermore, the more tree frogs we see, the less risk we inductivists run of making inferences that aren't good. For when a tree frog's chance of being green isn't close to 1, the more tree frogs we see, the less the chance that they'll all or nearly all be green and so induce us to infer—badly—that others are. That's what warrants the way in which our tendency to infer that tree frogs are green gets stronger as we see more and more tree frogs which are all or nearly all green.

That's how induction is warranted (and counter-induction unwarranted) by simple laws linking a pair of observable properties. But laws are rarely so simple. A frog's chance of being green may depend on many things. Hence the other factor in induction: the way in which our tendency to infer that frogs are green also increases with the diversity of the frogs which we've seen to be all or nearly all green.

Suppose you and I both see many frogs, but I only see tree frogs, while you also see many frogs of other kinds. How will this help you to make better inductive inferences about frogs? It won't if all frogs have the same chance of being green (or whatever): then we're on a par. But it will if they don't. For suppose tree frogs have a high chance of being green, and others don't. Then the inference 'it's a frog, so it's green' is good when it's a tree frog and bad when it's not. And because my chance of seeing only green frogs is now greater than yours (since I only see tree frogs), my chance of being induced to make this inference when it's bad is greater than yours. Whereas, provided you notice that tree frogs differ from other frogs, your chance of seeing only green tree frogs, and hence of being induced to make the good inference 'it's a tree frog, so it's green', is just as great as mine.

In short, whenever diversity in our inductive data matters, it's better to have it than not, and better to notice it than not. Of course we won't know if it matters at the time, since that will depend on laws we don't yet know. So the only warrantable tendency we can have is to be always readier to make inductive inferences when we know our data are diverse than when we know they aren't. And this is the tendency we do have.

But what if our unseen frog has no chance, high or low, of being green? In other words, no law, deterministic or statistical, simple or complex, links being a frog of whatever kind this one is with being green. Then there's no good inference from this being a frog of any such kind to its being green, or to its not being green. Neither of the inferential habits which induction or counter-induction might give us is good. But equally there's no specific chance, high or low, of induction or counter-induction giving us either habit, since there's no specific chance of many frogs of such a kind being all or nearly all green, or all or nearly all not green. In this lawless situation, neither induction nor counter-induction is warranted.

How do we know this isn't our situation? Maybe we don't know. But we don't need to know. Induction only fails here because here there's no success—no good habit of inference—to be had. When there is a good habit to be had, induction will always give us our best chance of getting it; and the better the inference, the better the chance. And that's enough to warrant induction as a general basic tendency.

But is it enough for induction to warrant the specific inference that this lecture is terse? Well, it is if lectures (or at least Inaugurals in English) have a high chance of being terse. For this will both warrant the habit of inferring that such lectures are terse and give induction a high chance of inducing this warranted habit, as it has done.

But do lectures like this have a high chance of being terse? Of course they do. You may not know this (though I dare say you do), but that doesn't matter. All that matters is that they do. And this fact (which, since you do at least believe it, you can't honestly deny) entails that your belief that this lecture is terse is both warranted, and warranted by induction. And so you do know already that it's terse—that it will last less than twenty-four hours—because, as you'll see at the end of this sentence, this belief of yours is not only warranted, it's true.

Reading 16.8 EXERCISE 9

From: Wittgenstein, L. (1969). *On Certainty*, Oxford: Basil Blackwell (Extract various paragraphs)

1. If you do know that *here is one hand*,[1] we'll grant you all the rest.
   When one says that such and such a proposition can't be proved, of course that does not mean that it can't be derived from other propositions; any proposition can be derived from other ones. But they may be no more certain than it is itself. (On this a curious remark by H. Newman.)

2. From its *seeming* to me—or to everyone— to be so, it doesn't follow that it *is* so.
   What we can ask is whether it can make sense to doubt it.

3. If e.g. someone says "I don' t know if there's a hand here" he might be told "Look closer".—This possibility of satisfying oneself is part of the language-game. Is one of its essential features.

4. "I know that I am a human being." In order to see how unclear the sense of this proposition is, consider its negation. At most it might be taken to mean "I know I have the organs of a human". (E.g. a brain which, after all, no one has ever yet seen.) But what about such a proposition as "I know I have a brain"? Can I doubt it? Grounds for *doubt* are lacking! Everything speaks in its favour, nothing against it. Nevertheless it is imaginable that my skull should turn out empty when it was operated on.

5. Whether a proposition can turn out false after all depends on what I make count as determinants for that proposition.

6. Now, can one enumerate what one knows (like Moore)? Straight off like that, I believe not.—For otherwise the expression "I know" gets misused. And through this misuse a queer and extremely important mental state seems to be revealed.

7. My life shews that I know or am certain that there is a chair over there, or a door, and so on.—I tell a friend e.g. "Take that chair over there", "Shut the door", etc. etc.

8. The difference between the concept of 'knowing' and the concept of 'being certain' isn't of any great importance at all, except where "I know" is meant to mean: I *can't* be wrong. In a law-court, for example, "I am certain" could replace "I know" in every piece of testimony. We might even imagine its being forbidden to say "I know" there. [A passage in *Wilhelm Meister*, where "You know" or "You knew" is used in the sense "You were certain", the facts being different from what he knew.]

9. Now do I, in the course of my life, make sure I know that here is a hand—my own hand, that is?

10. I know that a sick man is lying here? Nonsense! I am sitting at his bedside, I am looking attentively into his face.—So I don't know, then, that there is a sick man lying here? Neither the question nor the assertion makes sense. Any more than the assertion "I am here", which I might yet use at any moment, if suitable occasion presented itself.—Then is "2×2=4" nonsense in the same way, and not a proposition of arithmetic, apart from particular occasions? "2×2=4" is a true proposition of arithmetic—not "on particular occasions" nor "always"—but the spoken or written sentence "2×2=4" in Chinese might have a different meaning or be out and out nonsense, and from this is seen that it is only in use that the proposition has its sense. And "I know that there's a sick man lying here", used in an *unsuitable* situation, seems not to be nonsense but rather seems matter-of-course, only because one can fairly easily imagine a situation to fit it, and one thinks that the words "I know that . . ." are always in place where there is no doubt, and hence even where the expression of doubt would be unintelligible.

69. I should like to say: "If I am wrong about *this,* I have no guarantee that anything I say is true." But others won't say that about me, nor will I say it about other people.

70. For months I have lived at address A, I have read the name of the street and the number of the house countless times, have received countless letters here and given countless people the address. If I am wrong about it, the mistake is hardly less than if I were (wrongly) to believe I was writing Chinese and not German.

71. If my friend were to imagine one day that he had been living for a long time past in such and such a place, etc. etc., I should not call this a *mistake*, but rather a mental disturbance, perhaps a transient one.

72. Not every false belief of this sort is a mistake.

73. But what is the difference between mistake and mental disturbance? Or what is the difference between my treating it as a mistake and my treating it as mental disturbance?

74. Can we say: a *mistake* doesn't only have a cause, it also has a ground? I.e., roughly: when someone makes a mistake, this can be fitted into what he knows aright.

75. Would this be correct: If I merely believed wrongly that there is a table here in front of me, this might still be a mistake; but if I believe wrongly that I have seen this table, or one like it, every day for several months past, and have regularly used it, that isn' t a mistake?

76. Naturally, my aim must be to give the statements that one would like to make here, but cannot make significantly.

77. Perhaps I shall do a multiplication twice to make sure, or perhaps get someone else to work it over. But shall I work it over again twenty times, or get twenty people to go over it? And is that some sort of negligence? Would the certainty really be greater for being checked twenty times?

78. And can I give a *reason* why it isn't?

[1] See G. E. Moore, "Proof of an External World", *Proceedings of the British Academy*, Vol. XXV, 1939; also "A Defence of Common Sense" in *Contemporary British Philosophy, 2nd Series*, Ed. J. H. Muirhead, 1925. Both papers are in Moore's *Philosophical Papers*, London, George Allen and Unwin, 1959. *Editors.*

79. That I am a man and not a woman can be verified, but if I were to say I was a woman, and then tried to explain the error by saying I hadn't checked the statement, the explanation would not be accepted.

80. The *truth* of my statements is the test of my *understanding* of these statements.

81. That is to say: if I make certain false statements, it becomes uncertain whether I understand them.

82. What counts as an adequate test of a statement belongs to logic. It belongs to the description of the language-game.

83. The *truth* of certain empirical propositions belongs to our frame of reference.

84. Moore says he *knows* that the earth existed long before his birth. And put like that it seems to be a personal statement about him, even if it is in addition a statement about the physical world. Now it is philosophically uninteresting whether Moore knows this or that, but it is interesting that, and how, it can be known. If Moore had informed us that he knew the distance separating certain stars, we might conclude from that that he had made some special investigations, and we shall want to know what these were. But Moore chooses precisely a case in which we all seem to know the same as he, and without being able to say how. I believe e.g. that I know as much about this matter (the existence of the earth) as Moore does, and if he knows that it is as he says, then *I* know it too. For it isn't, either, as if he had arrived at his proposition by pursuing some line of thought which, while it is open to me, I have not in fact pursued.

85. And what goes into someone' s knowing this? Knowledge of history, say? He must know what it means to say: the earth has already existed for such and such a length of time. For not *any* intelligent adult must know that. We see men building and demolishing houses, and are led to ask: "How long has this house been here?" But how does one come on the idea of asking this about a mountain, for example? And have all men the notion of the earth as a *body*, which may come into being and pass away? Why shouldn't I think of the earth as flat, but extending without end in every direction (including depth)? But in that case one might still say "I know that this mountain existed long before my birth."—But suppose I met a man who didn't believe that?

86. Suppose I replaced Moore's "I know" by "I am of the unshakeable conviction"?

87. Can't an assertoric sentence, which was capable of functioning as an hypothesis, also be used as a foundation for research and action? I.e. can't it simply be isolated from doubt, though not according to any explicit rule? It simply gets assumed as a truism, never called in question, perhaps not even everformulated.

88. It may be for example that *all enquiry on our part* is set so as to exempt certain propositions from doubt, if they are ever formulated. They lie apart from the route travelled by enquiry.

89. One would like to say: "Everything speaks for, and nothing against the earth's having existed long before . . ."

Yet might I not believe the contrary after all? But the question is: What would the practical effects of this belief be?—Perhaps someone says: "That's not the point. A belief is what it is whether it has any practical effects or not." One thinks: It is the same adjustment of the human mind anyway.

93. The propositions presenting what Moore '*knows*' are all of such a kind that it is difficult to imagine *why* anyone should believe the contrary. E.g. the proposition that Moore has spent his whole life in close proximity to the earth.—Once more I can speak of myself here instead of speaking of Moore. What could induce me to believe the opposite? Either a memory, or having been told.—Everything that I have seen or heard gives me the conviction that no man has ever been far from the earth. Nothing in my picture of the world speaks in favour of the opposite.

94. But I did not get my picture of the world by satisfying myself of its correctness; nor do I have it because I am satisfied of its correctness. No: it is the inherited background against which I distinguish between true and false.

95. The propositions describing this world-picture might be part of a kind of mythology. And their role is like that of rules of a game; and the game can be learned purely practically, without learning any explicit rules.

96. It might be imagined that some propositions, of the form of empirical propositions, were hardened and functioned as channels for such empirical propositions as were not hardened but fluid; and that this relation altered with time, in that fluid propositions hardened, and hard ones became fluid.

97. The mythology may change back into a state of flux, the river-bed of thoughts may shift. But I distinguish between the movement of the waters on the river-bed and the shift of the bed itself; though there is not a sharp division of the one from the other.

98. But if someone were to say "So logic too is an empirical science" he would be wrong. Yet this is right: the same proposition may get treated at one time as something to test by experience, at another as a rule of testing.

99. And the bank of that river consists partly of hard rock, subject to no alteration or only to an imperceptible one, partly of sand, which now in one place now in another gets washed away, or deposited.

105. All testing, all confirmation and disconfirmation of a hypothesis takes place already within a system. And this system is not a more or less arbitrary and doubtful point of departure for all our arguments: no, it belongs to the essence of what we call an argument. The system is not so much the point of departure, as the element in which arguments have their life.