Alternative Approaches to Assessing the Effectiveness of Health Promotion Interventions

by

Simon Carroll

M.A., Lancaster University, 1997
B.A., University of Victoria, 1994

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

In Interdisciplinary Studies

© Simon Carroll, 2007
University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by photocopy or other means, without the permission of the author.
Alternative Approaches to Assessing the Effectiveness of Health Promotion Interventions

by

Simon Carroll

M.A., Lancaster University, 1997
B.A., University of Victoria, 1994

SUPERVISORY COMMITTEE

Dr. Marcia Hills, Supervisor, (School of Nursing)

Dr. Cecilia Benoit, Co-Supervisor (Sociology)

Dr. Irving Rootman (Faculty of Human and Social Development)

Dr. Marjorie MacDonald (School of Nursing)

Dr. David McQueen, External Examiner (Centers for Disease Control, Atlanta, USA)
ABSTRACT

Health promotion interventions are a key public health strategy aimed at improving population health and reducing health inequities. It is paramount that societies be able to assess whether investments in health promotion strategies are effective. However, the traditional or orthodox, evidence-based medicine approach to assessing the effectiveness of medical interventions has not proved to be appropriate for assessing health promotion interventions. Thus, there is a desperate need for workable, alternative approaches to assessing health promotion effectiveness. This study reviews current work in the field aimed at developing such alternatives and provides a critical analysis of some of the deep methodological challenges that confront this work.

Through the use of a specific case study of a project aimed at implementing one of these alternative approaches (‘realist synthesis’), this dissertation offers an original,
empirically based analysis of the types of emerging issues that one particular attempt at developing an alternative approach must deal with.

This study offers an ethnomethodological analysis of video-recorded 'work sessions', where the research team is working through specific aspects of the 'realist synthesis' project; particularly the conceptualization and categorization of 'contextual factors', and the identification and specification of 'outcomes'. Through this analysis, it is demonstrated that alternative approaches to synthesizing evidence for health promotion interventions must rely upon the mundane, practical, everyday competencies of 'abstraction' and 'objectification' as they are carried out, for all practical purposes, by a research team, as it goes about solving problems and completing normal research tasks.

The study concludes that much more attention needs to be paid to developing skills, representational tools and training research assistants, if alternative approaches are to be successful in completing reviews and sustaining credibility for end-users.
DEDICATION

This dissertation is dedicated to my parents, Derek and Agatha. Without their belief and support, this would not ever have come into being. I have been blessed with two wonderful people as parents, and for that I cannot be grateful enough.
ACKNOWLEDGEMENTS

The first acknowledgements are, as is appropriate for a doctoral dissertation, academic. My primary debt of gratitude goes to my supervisor, Marcia Hills. She has been my greatest supporter and has offered me more guidance and encouragement than I probably deserved. Marcia has taught me much about community based research, and has introduced me to the world of health promotion in her own imitable fashion: with pace, enthusiasm and unsurpassed dedication. She has fulfilled her role as both friend and mentor better than any student could have expected; she is, quite simply, an exemplary supervisor. In addition, I would like to thank my supervisory committee for their support and encouragement as I have wound my way towards this goal.

I have many, many other people who have had an influence on my academic progress, such that it is; however, some deserve a specific mention. At Lancaster University, where I completed my Masters degree, Bob Jessop, Andrew Sayer and John Hughes had a profound influence on my thinking, particularly about the philosophy of social science. Bob and Andrew are amongst the most lucid and persuasive expositors of critical realism, while John represents ethnomethodology, with its no holds barred approach to sociology, at its very best. I would be remiss not to mention the great support of my fellow graduate students over the years, particularly, Gord Miller at the University of Victoria, and Tiago Moreira, Majid Yar and Vince Miller at Lancaster.

Finally, last but not least, I want to thank my wife, Lora. There is a reason that spouses of PhD students get the tag, ‘long suffering’; however, with this patience and endurance goes an equal share in accomplishment. I literally, could not have done it without her love and support.
TABLE OF CONTENTS

Abstract ........................................................................................................ iii
Dedication ....................................................................................................... v
Acknowledgements ......................................................................................... vi
Table of Contents ........................................................................................ vii
LIST OF FIGURES ........................................................................................ viii

Chapter 1 An Introduction to Health Promotion Effectiveness ................. 1
The Values of ‘Scientism’ versus a ‘Science’ of Values? ................. 1
Health Promotion Effectiveness in the Context of Public Health ........ 2
Politics and Philosophy in Public Health and Epidemiology .......... 7
Governmentality and the ‘Trust in Numbers’ ...................................... 17
Summary ...................................................................................................... 21
Study Purpose .............................................................................................. 22

Chapter 2: A Review of the Literature: Substantive and Methodological .... 27
Introduction ................................................................................................. 27
Section 1: Substantive Literature on Health Promotion Effectiveness .... 29
Health Promotion: New Discipline or Multi-Discipline? ................. 29
Complexity ................................................................................................. 34
Section 2: Methodological Troubles ......................................................... 53
Research Question ....................................................................................... 74

Chapter 3 Methodology and Methods: ....................................................... 77
An Ethnomethodological Approach to the Sociology of Social Scientific Knowledge ................................................................. 77
‘Overturning’ Social Science ‘Indifferently’: Ethnomethodology as a Radical Critique of ‘Constructive Analysis’ ......................................................... 79
Methods and Data Collection ..................................................................... 91
Participant Observation ............................................................................. 91

Chapter 4 Results and Discussion ................................................................. 93
Categorization, Abstraction and the Situated Conceptualization of Context ................................................................. 93
The Ineluctable Ambiguity of Outcomes ............................................... 112
Summary ..................................................................................................... 124

Chapter 5 Conclusion and Recommendations ........................................... 127
Health Promotion Effectiveness ............................................................... 127
Theoretical Conclusions and Recommendations ............................. 129

References .................................................................................................. 135
LIST OF FIGURES

Figure 1    ‘The work setting’  93
CHAPTER 1

AN INTRODUCTION TO HEALTH PROMOTION EFFECTIVENESS

The Values of ‘Scientism’ versus a ‘Science’ of Values?

In order to understand the tensions within the field of health promotion, in relation to the question of how to evaluate health promotion effectiveness, it is necessary to first consider the relationship between the history of public health, the science of epidemiology and health promotion. As McQueen (2001) has pointed out, the ‘evidence debate’ is a precarious arena for health promotion to play in, and yet, it can hardly avoid participating. Health promotion, if it is serious about defending its values, approaches and methods for improving population health, must face squarely the ‘larger context’ of evidence-based medicine and other important influences and pressures on health promotion to demonstrate its effectiveness (McQueen, 2001; Nutbeam, 1999; Rootman, Goodstadt, Potvin & Springett, 2001). It must understand better the philosophical underpinnings of the ‘evidence-based’ movement, but just as importantly, it must come to grips with its political and social power and how this power is the result of complex historical forces, which have, if anything, increased and not diminished their influence. Health promotion should be especially wary about counting on the warm breeze of ‘post-modernism’ to come to its rescue (O’Neill, Rootman & Pederson, 1994); in fact, it is arguable that, rather than entering an age of ‘post-modernity’, we are in the midst of the triumphant crescendo of an elaborate hyper-modernity, with a level and tempo of rationalization that even Weber himself could not have envisaged. It is particularly dangerous for health promotion to wrap itself in the post-modernist flag, because at its core, health promotion depends on a dialectical and critical modernism that ceaselessly strives to fashion something valuable out of the ‘maelstrom of modern life’
(Berman, 1982). The question then becomes, how does health promotion face this challenge of the evidence debate, while retaining a commitment to its core values of public participation and the empowerment of individuals and communities to control the determinants of their own health? The purpose of this chapter is to explicate the tensions within the health promotion debate on how to assess or evaluate effectiveness, especially in relation to the science of epidemiology and the history of public health in the context of the wider social and political development of modern society. These tensions will be explored in three major sections: 1) the challenge of demonstrating health promotion effectiveness in the context of public health; 2) a historical analysis of the origins and philosophical and political underpinnings of public health and epidemiology; and 3) an interpretation of this history in relation to two key theoretical insights developed by analyses of the concept of ‘governementality’ (Burchell, Gordon and Miller, 1991; Dean, 1999; Foucault, 2000 [1978]) and the philosophical, historical and sociological analyses of the development of the statistical and probability sciences (Hacking, 1990; Porter, 1995).

Health Promotion Effectiveness in the Context of Public Health

Many health promotion researchers have prefaced their scholarly remarks on the birth and development of health promotion with a discussion of the sometimes-tortured relationship of the field to its older, more developed, paternal discipline, public health (Kickbusch, 1986; Terris, 1992). How this history is understood is perhaps the most telling aspect of how health promotion and its progress are viewed as a contemporary phenomenon and how the challenge and demand for scientific evidence of effectiveness is conceptualized and understood. The argument to be developed below is that there has been a tendency within health promotion to tell a story of public health as a ‘fall from grace’, from its original
reforming, perhaps even zealous, focus on the social and environmental causes of ill health, to a more restrictive preventive medical era, and finally to broader scale but narrower scope in the ‘lifestyles’ approach focused on individual risk factors and behavioural change (Kickbusch, 1986). In medias res, health promotion steps into the story to herald the era of a ‘new public health’, as a sort of re-emergence of the spirit of the 19th century socio-environmental model, with a modern gloss on the more subtle socio-economic determinants of health. The purpose of this critical analysis is to challenge this tendency to nostalgia, and to explicate some of its continuing consequences for health promotion’s rather schizophrenic relationship to public health. The ‘golden age’ of public health was primordially influenced by a particular philosophical and political outlook that still finds its expression today in its most modern and ‘rigorous’ proponents. If health promotion is to reinvigorate its relationship with public health it must be vigilant and cognizant of the fact that the power of the epidemiologists within public health is not an unfortunate deviation in the history of the field which is likely to go away anytime soon; this ‘power’ has to be analyzed and critiqued, if it is to be properly and fairly confronted, for confrontations are surely going to continue (Davey-Smith, Ebrahim & Egger, 2000).

The history of health promotion conventionally begins with the publication of the ‘Lalonde Report’, entitled A new perspective on the health of Canadians (1974). The report was notable because it was the first high level national government document anywhere in the world to advocate for health promotion as a basic strategy for improving population health. It was very influential internationally and it set the stage for future debate with its concept of the health field as the articulation of the argument that the medically dominated

---

1 Although it is just as conventionally noted that ‘health promotion’ activities, in one way or another, in various cultures, have been around since time immemorial.
health care system was only one and perhaps the least significant determinant of health, alongside biology, the physical and social environment and individual ‘lifestyles’. The Lalonde Report relied explicitly for its argumentation on such critiques of the health care system found in ‘social medicine’, as those comprehensively outlined by McKeown (1971, 1976), but which had their roots in the classic public health tradition of William Petty, Johann Frank, Rudolf Virchow and William Farr (White, 1991). In the Report, there is not only the notable (and much commented on) tension between an emphasis on individual ‘lifestyles’ and the subsequently neglected ‘socio-environmental’ factors, but an equal tension, given its own chapter heading as ‘Science versus Health Promotion’. Here it is made very clear that the ‘science base’ of the health field concept is epidemiology, and in this context, health promotion is seen as that type of action that must be taken, even though the pertinent scientific questions have yet to be definitively answered. In some ways this attitude allowed some initial breathing space for health promotion to get off the ground, but by setting up this dichotomy, it ensured that when the bets were finally called in and ‘science’ made its accounting, health promotion would have its day of reckoning with epidemiology.

Meanwhile, many in the health promotion community, especially in Europe and Canada, were starting to develop an independent conceptual basis for their work, based on a rigorous reflection on the type of actions necessary to most effectively promote the health of individuals and communities. Much of this work evolved out of a complex internal critique of the failure of traditional health education approaches, and a more sophisticated understanding of behavioural change (Kickbusch, 1986; Labonte, 1994; Tones, 1993). However, in Canada, the absence of a strong health education tradition was likely one of the factors which allowed it to contribute strongly to a more socio-ecological approach to
promoting health (with some its most influential leaders being sociologists and nurses, rather than psychologists) (Green, 2001); furthermore, for a variety of complex reasons (including again, individual leadership) much of the discourse of contemporary social movements (new leftist, feminist, gay/lesbian, environmentalist) found its way onto the official agenda of major institutions such as the World Health Organization (WHO) and Health and Welfare Canada (Labonte, 1994). As has been recognized by one of the leaders in health promotion internationally, Canada provided a hybrid and fertile mixture of traditional welfare state values and innovative community activism that seemed to provide the perfect ground for a push for the new socio-environmental approach to health promotion (Kickbusch, 1994). Out of this productive interaction between European ‘health promotion tourists’ (Kickbusch, 1986, 1994) and many able Canadian activist/public health practitioners grew the idea and finally the accomplishment of the Ottawa Charter for Health Promotion (1986) and the ‘Epp Framework’ (1986).

In order to fully understand the impact that the Charter had and continues to have, in relation to the specific question of the effectiveness of health promotion, it is important to see that there was a crucial transformation from the epidemiological and bureaucratic dominance of the Lalonde Report to the emerging “more pluralistic (and messier) social-science paradigm of human and social relations” embedded in the Charter (Labonte, 1994, p. 86). This shift is key to understanding the constant tension between a ‘scientific’ approach to evaluation and the ‘values’ underlying the Charter that is renewed each occasion that health promoters are asked to more rigorously account for their activities. It raises the uncomfortable question for those who, correctly, see the importance of reconnecting health promotion to the new public health, of just how ‘new’ the new public health is willing to be,
when it comes to its underlying philosophical commitments, especially its ontology, epistemology, and last but not least, its ethics.

This same tension underlies some of the confusion within the Canadian health promotion research community about how to relate to the emphasis on 'population health' as advocated by the Canadian Institute for Advanced Research (Coburn & Poland, 1996; Labonte, 1995, 1997; Poland et al., 1998; Raphael and Bryant, 2000, 2002; Robertson, 1998; Wong, 1997). On the one hand there is a justified admiration for the advocates of population health for their influential arguments about socio-economic determinants of health: even so far as to single out 'progressive' population health researchers such as John Frank (Raphael & Bryant, 2002); while, on the other hand there is the more or less well-articulated angst about the lack of health promotion principles within the population health perspective. The critiques of the population health perspective for its lack of emphasis on values, it weak or non-existent orientation to action, and its somewhat imperious attitude to what is to count as proper 'knowledge', are all cogent and well aimed. The question is: why should this be a surprise? It isn't enough to point out and lament the baleful influence of a replacement ideology for health promotion. Where did it come from, and why is it so influential?

Furthermore, is this newly emergent approach (Evans & Stoddard, 1990) really so new? How far is it simply a modern, sophisticated renaissance of that very same 'golden era' of public health that health promoters so often return to as their intellectual and moral heritage.

Although health promoters themselves (especially Canadians, who were the ones facing this challenge directly) reacted strongly and were able to defend the rationale for keeping a health promotion focus, the more incisive critiques were often too 'reactionary'
and came off rhetorically as overly defensive. A more ‘accommodating’ response came from within Health Canada itself (Hamilton & Bhatti, 1996).

Why is it that health promotion often seems ‘behind the game’ in the science debate? Partly, this is due to the fact that, as Labonte says, the social science world is ‘messy’, as a recent Companion to Social Theory attests (2000)²; however, may it also be partially true that for too long health promotion has neglected its need to develop an independent ‘science base’ having unconsciously bought into that original Lalonde dichotomy? This is becoming increasingly clearer as many of the leading proponents in the field are pushing for more intensive theoretical development, a pressure that has become especially acute as the need for demonstrating effectiveness has increased (McQueen, 2001; McQueen & Jones, 2007; Nutbeam, 1999). To understand why health promotion has such a complex and ambiguous relationship to public health, it is necessary to dig more deeply into the foundations of modern public health and to unpack its driving philosophy and world-view.

Politics and Philosophy in Public Health and Epidemiology

To begin, it is crucial to understand that the roots of the modern public health epidemiologists’ focus on individual risk factors and randomized controlled trials (RCTs) is not in contradiction with or a deviation from the Edwin Chadwicks, the John Simons and the John Snows of the classical public health; rather, the full flowering of individualistic, utilitarian calculus, an uncompromising economism, and an obdurate scepticism of anything but positivistic scientific knowledge, can be seen as the late fruit of more than three centuries of development in public health and epidemiology. If health promotion really is to engage in

² Social science is ‘messy’ both because its phenomena of interest are enormously complex, but also because the various disciplines themselves are internally fractured in terms of philosophical and methodological approaches, particularly contemporary sociology.
a dialectical approach to the individual and society, health and illness, quantitative and qualitative methods, and reductionistic and holistic world-views, it must confront squarely the truth that its concerned *pater familias* in the discipline of public health carries a lot of baggage with it.

The first part of the argument sets out the dilemma and *prime facie* paradox that health promotion faces when confronting its genealogy in the history of public health. In terms of lives saved and healthy years lived, the early public health interventions to combat the spread of deadly and debilitating communicable diseases cannot be underestimated. However, it is no accident that once the environmental risk factors of the major communicable diseases were effectively neutralized, that a shift in focus took place to providing preventive, immunization measures. As Kerr White (1991) so convincingly puts it, the history of public health and epidemiology can be read as successive and iterative ‘redefinings of the unacceptable’. Public health has always been concerned with an economistic and utilitarian approach to the health of the population; when things start to ‘cost too much’, the unacceptable becomes miraculously ‘visible’. To understand this history, one has to ignore the facile disciplinary chasm between public health and economics: a chasm which has only recently and tentatively been bridged (Evans, Barer & Marmor, 1994); the fact is that while economics became progressively theoretical and mathematical, public health continued the original classical liberal tradition of reformist, practical utilitarianism, most powerfully apparent in the Benthamite tradition’s attempt to rationalize government and public services. There is a great irony that the humanitarian idealism (an idealism at the core of health promotion’s values base) of the British ‘public health doctors’, such as Haygarth, Heysham, Thackrah, Baker and Millar, was never the driving force behind concentrated
public health action (Fraser, 1973). The purpose here is not to simply undermine such a legacy. The complexity that attends this history is part and parcel with the whole legacy of modernity and modernism, the state and all its various accoutrements. It is a history that incorporates the contradictory tendencies elaborated so succinctly by the great modernist thinkers: from Marx to Freud to Nietzsche; from Weber to Simmel to Durkheim; from Goethe to Dostoyevsky to Kafka (Berman, 1982). We can no more reject the trust in numbers, positivism, rationalization, and economism, than we can reject wholly the more Romantic aspects of modernism, with its distrust of rationalization and scientism, and its penchant for a nostalgic re-animation of community, spirituality and wholeness.

Health promotion often finds itself as an unconscious victim of this dialectic, unable to place itself in relation to the powerful historical forces that have driven the agenda of public health over the past two centuries. Health promotion, as it has evolved, is condemned to navigate between the Scylla of a state-centred bureaucratic logic of rationalization and the Charybdis of a nostalgic and largely impotent ‘community’ romanticism. Neither fate is necessary, but the science and evidence debate is perhaps the most severe test for health promotion’s navigation skills. In what follows, there is an attempt to more directly unpack some of the subtleties and ambiguities underlying this dilemma.

We follow here a two-sided story of the unfolding of a specific and influential scientific discipline, and across this warp, the weft of the slow institutionalisation and hesitant development of an attitude to government, that is only now starting to express, fully, its powerful design. An underlying irony is how the formation and progress of ‘political arithmetic’, which originated in the 17th Century, only found its most excessive employment at precisely the time when the end of modernity was being heralded? By reflecting on public
health, epidemiology and its subsequent development, some core philosophical themes are examined, themes which reflect both on the origins and purposes of public health, as well as on its future prospects and its impact on health promotion.

‘Public health’ refers to a movement of reform that involved an effort of ‘state-building’, receiving its first serious impetus in Britain with the Poor Law Amendment Act of 1834. This was a combination of civil engineering, preventive health care and local administrative action, which combined to advocate a more coherent strategy for ameliorating the worst of rapid industrialization and urbanization’s excesses. However, as Derek Fraser (1973) points out, from the last quarter of the 18th Century onwards, local doctors were steadily reporting the link between environmental conditions, social context and the health of populations. These disparate reports culminated in the famous survey of Manchester by James Kay, entitled The Moral and Physical Condition of the Working Classes (1832). The most famous name in the annals of the public health movement was, Edwin Chadwick. In 1832, Chadwick, a committed Benthamite, who was in fact a secretary to Bentham himself, accepted an appointment on the Royal Commission on the Poor Law, and was the joint author of the report with Nassau Senior, the famous classical economist. The purpose of the Commission was to examine the working of the Poor Law. Chadwick quickly recognized that the principle of ‘less eligibility’ could not account for the major problem of whole families made destitute due to the main earner being incapacitated by ill health. Chadwick

---

3 By way of example, Britain is used a paradigm case, although varying manifestations of a similar story could be told about Continental Europe, particularly in Germany and France (White, 1991). Nevertheless, the Anglo-American development of public health and epidemiology has been the dominant force behind the positivistic aspect of public health evaluation of effectiveness and therefore justifies this somewhat narrow focus.

4 The emphasis in the Poor Law reforms that insisted on eligibility ‘tests’ to judge the fitness to work of the claimant to relief. The idea was that paid work must always be more ‘attractive’ than receiving hardship funds under the Poor Law; given that paid work, at that time, was often a hellish and impoverished drudgery, making
recognized that the affliction of the major endemic and epidemic diseases of the great urban conglomerations could not be blamed on the moral character of the poor. Being a hard-headed pragmatist, Chadwick cut through the ideology and argued for the ‘Sanitary Idea’. The ‘Sanitary Idea’ consisted of three strands: 1) a theory of causation. Chadwick believed in the prevailing view of his time, that the ‘foul air’ generated by the decomposition of organic matter and waste caused urban diseases. His researches confirmed this, in that the worst areas of unsanitary conditions also had the highest incidence of disease. 2) He proposed a remedy for the problem. This was to be a massive civil engineering program, aimed at flushing away the filth and waste in major urban centres with a modern drainage system, the basis of which would be the glazed earthenware pipe. This was the emergence of the water closet (WC) as the core unit of modern sewerage (Chave, 1984). 3) He had a strategy of implementation:

...that for the general promotion of the means necessary to prevent disease it would be good economy to appoint a district medical officer, independent of private practice, with the security of special qualifications, with responsibility to initiate sanitary measures and reclaim the execution of the law (quoted Chave, p.6)

Before moving on to consider more of the historical detail, it is important to emphasize Chadwick’s adherence to the utilitarian creed, a creed that leaves a profound legacy to the tradition of public health, and one which reverberates throughout the entire discourse surrounding evidence-based medicine. Chadwick was not primarily concerned with the well-being of individual people, except in so far as their illnesses constituted unnecessary costs to the Poor Law system. Of course, utilitarianism was concerned with increasing

---

Poor Law relief seem ‘less attractive’ was a most sadistic accomplishment---cue the infamous workhouses, referred to by the workers as the ‘Bastilles’.

5 Not the first case of ‘confirmation’ of a causal claim being confused with a correlation of variables.
‘happiness’, but in an overwhelmingly economistic sense of the term, and Chadwick’s main
drive was, as a typical Benthamite reformer, to apply the ruthless logic of efficiency calculus
to a specific area of government. At its core, this injected a lasting push for generic
measurement devices that could aggregate definite properties of individuals into a corpus of
statistical data on the health of populations. As we shall see, public health found its scientific
fulcrum, and its specific logic of measurement and analysis of data, in the burgeoning
discipline of epidemiology. In fact, ‘public health’ as a discipline is really the
institutionalization of applied epidemiological studies, and epidemiology has always been, in
its essence, an applied science. Although, in its more modern version, the public health
movement is often associated with socialized medicine (Fraser, 1973), its roots and its
philosophical bearings are grounded solidly in the utilitarianism that it shares with orthodox
economic theory; itself deriving its core principles from classical, laissez-faire economics.
This is all commonplace in the historical debate about the growth of the administrative state
in the 19th Century Britain. Although some argue there were contradictions between the
laissez-faire convictions of Bentham and the paternalistic administrative zeal of the
Benthamites, the considered view is that the administrative reforms were always
implemented on the nose-holding principle that, if there were to be the dead hand of
government involved in the life of ‘free’ individuals, it should be ‘good’, ‘rational’, and
above all ‘efficient’.6 While public health reforms were certainly taken up and transformed
by socialists in the last century, the economistic bias always loomed in the background, and
the inherent drive towards a positivistic methodology of optimistic quantitative innovations
was set in place.

6 As A.J. Taylor put it: “...the Utilitarian contribution to the Welfare State, however real, was essentially an
unwitting one” (Taylor, 1972, 37).
One of the more patient and less irascible characters in the public health movement, but probably (next to Chadwick) one its most important figures, was a doctor named John Simon. Simon was appointed the first Medical Officer of Health (MOH) for London in 1848. The fact that he was a practicing clinician and not only an ‘interfering bureaucrat’, meant that where Chadwick managed to alienate, Simon managed to persuade. Both men found enemies in the local power bases that resented centralizing interference in local affairs, but Simon at least, was able to avoid annoying everyone simultaneously, and certainly convinced many, that action on public health reform was necessary. However, in the end, his lasting contribution was not the depth and breadth of improved sanitation (the course of progress in the 19th Century was very slow and at times stalled completely), it was in his patient advocacy of research and data collection that eventually influenced the establishment of an academic specialty in public health and thus left the legacy of a solid disciplinary foundation. This legacy has survived and prospered, despite its relative neglect in policy terms. This is crucial to understanding the present condition of public health. The legacy can be thought of in terms of two major influences: 1) the concern with the health of populations, tied to a basic utilitarian philosophy that prioritized a statistical methodology for comparisons of causal factors of disease and the most effective preventive measures to fight them. Thus, not only is there a focus on tracing the causes of disease in populations, but a consistent push to criticize those curative and preventive measures that are less cost-effective, and to advocate measures that are more cost-effective. 2) The establishment of an intellectual discipline as a cocoon for the growing sophistication of the science of epidemiology. It is to this latter subject that I now turn.

7 The Medical Officers of Health were instituted by Chadwick, crucially (as one can see from the quotation above) on the basis of a specialised skill in research in the field of preventive medicine.
‘Epidemiology’, literally “the study of that which is upon the people” has a long history. Its roots are in the origins of health statistics and ‘political arithmetic’, developments that were firmly grounded in the 17th Century Scientific Revolution, centred on The Royal Society. The major figure here, and the so-called ‘father’ of medical statistics is John Graunt (1620-1674), who published the famous treatise, Natural and Political Observations Mentioned in a Following Index and Made Upon the Bills of Mortality, in 1662. Graunt is credited with this title for his original contribution to an analytical approach to the Bills of Mortality. Undoubtedly, Graunt belonged to that generation of ‘gentlemen’ who were wholly committed to the notion of ‘Improvement’ (Porter, 2000). Graunt established through analysis of the mortality figures four basic phenomena of the population’s health characteristics: 1) he showed the regularity and pattern of vital phenomena at the level of the population. 2) He noted the excess of male births as opposed to female births. 3) He found a high rate of mortality in early life. 4) He discovered that the urban death rate was much higher than the rural death rate.

Even more important than Graunt was William Petty (1623-1687). Petty’s great conceptual contribution was the coining and elaboration of the term: Political Arithmetic. Petty was trained in medicine, received a Doctor of Physic degree from Oxford, and was appointed Professor of Anatomy at the tender age of 28. He was a prolific writer, and along with being one of the founders of the discipline of economics, he wrote on the subject of the health of populations. He was the first to systematically build the methods of statistical

---

8 It was these patterns or regularities, that became the focus of probability mathematics and statistical methodology, in order to establish whether such regularity could have occurred simply by chance, or whether, in fact, there was good reason to believe that they were due to underlying causal relations.
analysis into the study of populations. In his posthumous work, Political Arithmetic, Petty outlined concisely his approach:

The method I take is not yet very usual; for instead of using only comparative and superlative words, and intellectual arguments, I have taken the course [as a specimen of the political arithmetick I have long aimed at] to express myself in terms of number, weight, or measure; to use only arguments of sense, and to consider only such causes, as have visible foundations in nature; leaving those that depend upon the mutable minds, opinions, appetites, and passions of particular men, to the consideration of others (quoted in White, 1991, pp.31-32).

It is uncanny how contemporary a statement this feels and sounds. Indeed, Petty was extraordinarily ahead of his time, having "recommended a central statistical office 150 years before the General Register Office was established in Britain" (White, 1991, p. 31). Two other specific and prescient innovations by Petty were: the first serious attempt to assign comparable values to individual lives, anticipating the development of life tables and actuarial science; second, he inaugurated the first analytical scepticism about the benefits of medicine, and developed the embryonic arguments that were to form the basis of 'evaluative research' as they are practiced and proselytized at the heart of evidence-based medicine today. As White argues, after Petty, the face of clinical medicine and public health changed forever, and "introduced quantification to link what later became three related branches of learning: economics, epidemiology, and political science" (p. 32).

The three great innovations necessary for the full implementation of Graunt and Petty's vision, and the constitution of the science of epidemiology were: the actuarial science and life tables developed by the astronomer, Edmund Halley (1656-1742); the theory of probability, initiated by Condorcet (1743-1794); and the beginnings of collection of analysis of the mass survey (White, 1991). The actual use of the word 'statistik' comes from the
German physician and scholar in political theory, Herman Conring⁹, but acquired its modern conceptual form in the work of John Sinclair, a Scottish agricultural economist. However, the real innovation was the development of the theory of probability, with Pierre-Simon Laplace being one of the more crucial figures for epidemiology. In a primitive form, many of the essential elements of the logic of inferential statistics, including the basis for the future concepts of 'efficacy', 'case-control study', and 'randomization' can be found in Laplace. Following Laplace, Gauss and Poisson, greatly advanced the theory of probability, based on the work of the great 17th century mathematical philosophers: Bernoulli, Pascal, and Fermat.

It is worthwhile summarizing what this short history shows. We find in the people listed here a typical combination of the intellectual innovation of the ‘gentleman’ scholar, with the earnest concern for societal improvement of the ‘man of affairs’. Epidemiology is the apotheosis of this alchemical mixture. There is a general culture of inquiry, located amongst the ‘free’ post-Renaissance intellectuals and out of this culture gestates a new and optimistic attitude towards the possibility of ‘good government’, guided by the visible hand of reason.¹⁰ There are aspects to this culture and this attitude which are apparent in the hyper-modern drive to implement a rigorous ‘science’ of evaluation by way of an ‘evidence-based’ methodology. It is crucial to note that much of the attention to quantitative precision had less to do with fundamental science and much more to do with the construction of ‘objectivity’ as a cultural solution to the governmentality of the state. Next will be considered two strands of argument that approach this issue from quite different, though complementary perspectives.

---

⁹ Derived from the German for ‘state’, Staat.
¹⁰ This characterization is indebted to the late Roy Porter’s unparalleled contribution to the study of this peculiar period and its relation to Enlightenment values (Porter, 2000). Sometimes we are simply privileged to have been exposed to the life work of very special scholars; Roy Porter was certainly one of these.
Governmentality and the ‘Trust in Numbers’

Here it will be argued, based on the later work of Michel Foucault on ‘governmentality’, that the development of public health should be seen within a much broader setting of the flowering of various technologies of government and governance related to coping with the complexity of the modern state. This brief review of the work on ‘governmentality’ will be tempered by placing it the context of Theodore Porter’s work on the historical development of statistics and the cultural and social dominance of ‘numbers’ in modern societies.

The literature on ‘governmentality’, as it has been espoused by a variety of writers inspired by Foucault (Burchell et al, 1991; Dean, 1991; Osborne, 1993; Rose, 1993), is important because it offers a potentially fruitful way of understanding some of the emerging forms of political governance in advanced capitalist societies, without reducing them to immediate ideological dispositions or class allegiances. The concept of ‘governmentality’ has been used in several ways, in relation to recent changes in the nature of government in advanced capitalist societies. In particular, it has been used to analyse the overall context of governing the population’s health, and how such a ‘purpose’ arose and has been transformed (Foucault, 2000); it has also been used to analyse the relationship between the state and the medical professions (Osborne, 1993). These objects of analysis are not easily separable, and how they are understood determines much of the emphasis that results from applying the concept of governmentality to the area of contemporary changes in health care systems. The problem of ‘population’ as articulated by Foucault is at the centre of his analysis of ‘governmentality’; however, the theme of ‘population’ arises earlier in Foucault’s work in the first volume of *The History of Sexuality*, where the concept of the “bio-politics of the
population" is first articulated as one of "the two basic forms" of power over life as it evolved, starting at the end of the seventeenth century (Foucault, 1978). It is even more apparent in a paper where Foucault discusses what he terms the 'noso-politics' of health in the eighteenth century (Foucault, 1980). It in this latter paper where Foucault explicitly links the development of public health technologies with the themes of Polizeiwissenschaft or the 'science of policing'. Thus, at the very core of the concept of 'governmentality' are the technologies of 'public health', a 'noso-politics' that necessitates a vast array of 'vital statistics' concerning the population as an object of knowledge. There are a variety of authors who have followed up this analysis (Castel, 1991; Osborne, 1993). There is some disagreement about how to characterize the subsequent analysis Foucault made of 'liberalism' as a mode of governmentality that develops as a critique of state reason. Some emphasize 'liberalism' as a reaction to the mercantilist/cameralist 'police states', where the function of Polizeiwissenschaft is replaced by liberal forms of governance, while others emphasize the continuity and pluralism of various types of governing technologies. The relationship between continuity and discontinuity is a central motif in all of Foucault's work, and is a constant object of reflection in his various writings and interviews (see Foucault, 2000, p. 219). 'Governmentality' is simply the name Foucault gives to the overall context for the project of political governance since the emergence of 'population' as its principle target. 'Liberalism' is the term, which Foucault uses to distinguish the 'attitude' of Polizeiwissenschaft from that of a more circumspect and reticent mode of governance that

---

11 This article was originally published in French in 1976 as 'La politique de la sante' au XVIIIe siecle', in Foucault et al., Les Machines a Guerir (aux origines de l'hoptial moderne) (Institut de l'Environnement, Paris, 1976).
constantly questions whether ‘government’ is necessary at all. For Foucault, the German tradition of Polizeiwissenschaft is concerned with how to govern most effectively and efficiently in each ‘policy’ area; the problem to be solved is one of ‘under-governance’. Conversely, liberalism inaugurates a series of reflections that ask whether it is really necessary to ‘govern’ in the first place; here the issue is one of ‘over-governance’ (Foucault, 1994a, 73-79). This is especially difficult in relation to the discussion of public health. Here there is a structural tension between a ‘liberal’, economizing technology of governing the health of the population, and a more interventionist ‘policing’ of the population. For our purposes, it is important to locate the waxing and waning of the relevance and power of the science of epidemiology and its relation to efficient and publicly accountable administration within this complex architecture of governance in advanced modern, capitalist societies. The social forces within and outside the modern capitalist state are constantly trying to adjust to the contradictory demands and strategic dilemmas inherent in stitching together the needs of continuing capital accumulation and the necessity of guaranteeing a general social cohesion in representational, democratic societies (Jessop, 1992; 2000). The state can’t help but intervene in the fissures in the societal fabric generated by the ‘creative destruction’ of capitalist dynamics; yet, at the same time these interventions are subject to the pursuit of efficiencies in the modes and operations of government. To understand more clearly how these tensions relate to the health promotion effectiveness debate, we have to turn to the related but much more historically concrete analyses of the emergence of the social technologies of public accountability.

A variety of authors have contributed to the emerging work on the historical and sociological analysis of the broad cultural development of the ‘trust in numbers’ or as others
of called it, the ‘values of precision’ (Hacking, 1991; Porter, 1995). For some, such as Hacking (1991), there is an explicit connection between studies of governmentality and the historiography of statistical science. For others, such as Porter (1995), the connection is there, but the growth of the importance of ‘numbers’ and quantification is part of more general ‘cultures of objectivity’, particularly in the tension between ‘disciplinary’ cultures of objectivity and ‘mechanical’ ones. By ‘disciplinary objectivity’, Porter refers to a realm of objectivity “made conspicuous mainly by its absence”, with a prime example being the ‘large territory’ on which physicists are not “called upon by outsiders to justify their conclusions” (p. 4). In contrast, ‘mechanical objectivity’ develops as a reaction of those scientific and other ‘communities’ subject to an intrusive pressure to make their activities transparent and publicly accountable. The main driving force behind the rapid expansion of statistical inferential techniques and epidemiology itself has been the relative weakness of these communities:

On the whole, statistical inference has not made its way down the hierarchy of science, from mathematics and physics to the biological and at last the social sciences. Rather, it was seized most readily by weaker disciplines, such as psychology and medical research, and indeed by their relatively applied sub-fields (Porter, 1995, p. 200).

Thus, ironically, the push for health promotion to be more ‘scientific’ comes not from the established natural sciences or mathematics, both of whom look somewhat askance at the more florid innovations of applied statistics, but from the reflexive backlash of their equally weak and youthful competitors in the socio-medical and psychological realms. The important argument to keep in mind that is made by Porter, is that the drive for measurement precision and quantitative objectivity is less about a scientific realist’s pursuit of truth and more about the relationship between uncertain bureaucratic power in need of justifying its decisions and uncertain professional communities attempting to defend their autonomy and trustworthiness.
What this points to is that health promotion is not alone in its dilemmas in relation to pressures to evaluate effectiveness; there is a more generic cultural and political dynamic which Porter outlines, that underlies the specific nature of the quandary health promotion finds itself in. In other words, although epidemiology manifests itself as the specific approach which health promotion must accommodate or confront, epidemiology itself derives its political and social power from a general cultural dialectic between ‘trust’ and ‘accountability’.

Summary

In summary, it has been argued that to understand the dominance of epidemiology within public health, and how health promotion has been subjected to its scientific gaze, one has to reflect on the fact that, even in the fading light of its gilded origins, public health was historically grounded in a bureaucratic and utilitarian logic that always stepped forward uncomfortably with its more idealistic and humanitarian impulses. The political and social power of ‘political arithmetic’, the ancestral heritage of epidemiological science, is part of a long history of the institutionalization of a ‘trust in numbers’ as a method of controlling risk in the context of the modern governance of populations, and epidemiology is perhaps only one of its most well developed offspring. For health promotion, a reflective understanding of its effectiveness should entail a rigorous confrontation with the core of the epidemiological project within the context of its wider embeddedness in this social and political context. Furthermore, it must consider alternative philosophical approaches that have already developed trenchant critiques of the underlying basis of the epidemiological and ‘evidence-based medicine’ projects.
Study Purpose

The purpose of this study is to examine the ‘practice’ of actually doing the work of developing and implementing one of these alternative methodological approaches to reviewing the evidence for the effectiveness of health promotion interventions. The realist synthesis approach is a nascent theoretical and methodological alternative, which has recently been implemented in a series of projects in order to produce systematic reviews of evidence relating to the effectiveness of community health promotion initiatives. This study offers an empirical perspective on the in situ practice of the implementation of one of these projects, as it was carried out by a research team by the Community Health Promotion Research Centre, at the University of Victoria, under the direction of Dr. Marcia Hills.

This project had a series of research objectives:

- to further develop and refine the methodology of realist synthesis for public health and health promotion interventions.

- to conduct a pilot systematic review using the realist synthesis approach focused on synthesizing the evidence for the effectiveness of community efforts that impact on the social determinants of health and health disparities.

- to assess the generalizability and adaptability of the realist synthesis approach for health promotion and public health.

- to assess the practical relevance and timeliness of the realist synthesis approach for informing the policy and planning process for health promotion and public health interventions.

The guidelines developed by Pawson et al. (2004) were used in the systematic review, using the realist synthesis approach. There are a series of steps review process which are shown in the table below with subsequent descriptions, timelines and methods provided:
<table>
<thead>
<tr>
<th>Define the scope of the review</th>
<th>Identify the question</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>- What is the nature and content of the intervention?</td>
</tr>
<tr>
<td></td>
<td>- What are the circumstances or context for its use?</td>
</tr>
<tr>
<td></td>
<td>- What are the policy intentions or objectives?</td>
</tr>
<tr>
<td></td>
<td>- What are the nature and form of its outcomes or impacts?</td>
</tr>
<tr>
<td></td>
<td>- Undertake exploratory searches to inform discussion with review commissioners/decision makers</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Clarify the purpose(s) of the review</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Theory integrity – does the intervention work as predicted?</td>
</tr>
<tr>
<td>- Theory adjudication – which theories about the intervention seem to fit best?</td>
</tr>
<tr>
<td>- Comparison – how does the intervention work in different settings, for different groups?</td>
</tr>
<tr>
<td>- Reality testing – how does the policy intent of the intervention translate into practice?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Find and articulate the programme theories</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Search for relevant theories in the literature</td>
</tr>
<tr>
<td>- Draw up 'long list' of programme theories</td>
</tr>
<tr>
<td>- Group; categorise or synthesise theories</td>
</tr>
<tr>
<td>- Design a theoretically based evaluative framework to be 'populated' with evidence</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Search for and appraise the evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Decide and define purposeful sampling strategy</td>
</tr>
<tr>
<td>- Define search sources, terms and methods to be used (including cited reference searching)</td>
</tr>
<tr>
<td>- Set the thresholds for stopping searching at saturation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Appraise the evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Test relevance – does the research address the theory under test?</td>
</tr>
<tr>
<td>- Test rigour – does the research support the conclusions drawn from it by the researchers or the reviewers?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Extract and synthesise findings</th>
<th>Extract the results</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Develop data extraction forms or templates</td>
<td></td>
</tr>
<tr>
<td>- Extract data to populate the evaluative framework with evidence</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Synthesise findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Compare and contrast findings from different studies</td>
</tr>
<tr>
<td>- Use findings from studies to address purpose(s) of review</td>
</tr>
<tr>
<td>- Seek both confirmatory and contradictory findings</td>
</tr>
<tr>
<td>- Refine programme theories in the light of evidence</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Draw conclusions and make recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Involve commissioners/decision makers in review of findings</td>
</tr>
<tr>
<td>- Draft and test out recommendations and conclusions based on findings with key stakeholders</td>
</tr>
<tr>
<td>- Disseminate review with findings, conclusions and recommendations</td>
</tr>
</tbody>
</table>

Below is a more detailed outline of the steps in the research process:

1) Define the Scope of the Review (Months 1-2)

   There are three key components when clarifying the scope of a realist review: identifying the review question, refining the purpose of the review, and articulating key theories to be inspected. These activities require ongoing and iterative dialogue with the persons/commissioners/reviewers who will deliver and/or develop the emerging theories. The purpose of a realist review is explanatory and there is no particular strategy or technique for defining the scope of a review, rather there is a review logic. The process of articulating the scope of a review acknowledges that interventions are complex and multiply embedded. The task at hand then is to prioritize the key theories that are to be explored.

2) Search and Appraise the Evidence (Months 2-6)
A realist synthesis approach to searching for evidence is a more iterative and interactive process than traditional systematic reviews and the searching process continuously evolves as understanding of the subject matter deepens. In realist reviews, a wide breadth of empirical studies drawn from various bodies of literature are regarded as important, the ‘searching’ stage spans widely across the systematic review timeframe, and purposive sampling strategies are employed which test specific theories and answer particular questions in an iterative, reflexive manner.

Test relevance and rigor typically serve as benchmarks of appraisal within traditional systematic reviews; a realist approach places emphasis on making sense of varied contributions and getting a mixed picture from various forms of research while weighing the contributions of each source of evidence. It is held that “the worth of studies is established in synthesis” (p. 23) not in any kind of pre-qualification exercise such as using only randomized control trials as valuable sources of evidence.

3) Extract and Synthesize Findings (Months 4-8)

Pawson describes the data extraction stage as the “core” of systematic review and as one which lacks a precisely corresponding stage in realist review. Realist reviewers may draw from diverse sources and employ multiple approaches to data extraction; no single or standard form or set of questions is considered sufficient to the task of apprehending and consolidating relevant findings. Pawson recommends practices such as annotation and description of studies’ relative usefulness as alternatives to the conventional extraction methodologies of grid entries and matrix construction.

Pawson identifies the steps of realist review as non-linear, and posits the processes involved in extraction and synthesis as interrelated. He describes an “ill-defined point in the
sifting and sorting of primary models where one changes from framework building to framework testing and from theory construction to theory refinement” (p. 24). A comprehensive approach to data extraction is thus the foundation of theoretical synthesis. Pawson characterizes interventions as theories, and synthesis of evidence as a process of refining interventions/theories. He acknowledges the limitations of the reviewer in this endeavour; interventions are meant to be implemented, and part of the task of their refinement inevitably rests with those responsible for putting them into practice. Pawson delineates four different "cuts" or "slants" that realist synthesis can take. Common to each of these is an analytic focus on programme theory rather than primary studies.

4) Draw Conclusions and Make Recommendations

For valuable recommendations to be made, it is necessary that within the research synthesis, the practical needs of stakeholders are taken into account. Conclusions and recommendations that speak to issues that are on policy- and decision-makers agendas and that are illustrated in decision-making ‘language’ (Pawson, Greenhalgh, Harvey & Walshe, 2004) have a greater impact on knowledge translation. Involving stakeholders in the process of drawing conclusions and making recommendations can ensure that revisions to the initial understanding of an intervention are made.

A realist review provides an assessment of how an intervention or implementation works through existing research. Dissemination of conclusions will inform programs of research and enable individuals to account for complex and subtle findings and apply them to their local context. Conclusions and recommendations within a realist framework are not designed to serve as a final judgment on what works or not, rather the aim is to increase the understanding of how interventions are thought to work, resulting in a more refined theory.
This study, which is a reflective analysis of the work of the project outlined above, aims to contribute to the ongoing development of alternative approaches to assessing health promotion effectiveness that has been stimulated by the International Union for Health Promotion and Education’s ‘Global Programme on Health Promotion Effectiveness’. Its original contribution is to focus not on the theoretical debates and discussions that have recently been given impetus (McQueen & Jones, 2007; McQueen & Kickbusch, 2007), as necessary as these are; rather, it is to give a first-hand account, in a detailed empirical analysis, of the everyday challenges of implementing a realist synthesis approach. It follows the maxim that at some point an idea has to get off the couch and walk out the door before we can really see if it has legs.

In the next chapter, the theoretical and methodological issues that trouble the health promotion effectiveness enterprise are critically reviewed, before moving to the specific methodology used for this empirical study, the results and discussion, and finally, the conclusion and recommendations.
CHAPTER 2:

A REVIEW OF THE LITERATURE: SUBSTANTIVE AND METHODOLOGICAL

Introduction

This chapter falls into two sections: the first will provide a critical review of the substantive literature on health promotion effectiveness over the past two decades; the second section will examine certain broad methodological 'troubles' that persist in a variety of approaches to assessing the effectiveness of health promotion.

The first section focuses mainly upon a series of collective attempts to bring scientific credibility to the field of health promotion through collaborative examinations of the bases for evaluating the effectiveness of health promotion programs, interventions and polices. This runs somewhat counter to the normal focus of a critical literature review, which alights mainly upon key individual contributions\(^{12}\) to the field in terms of specific positions and arguments. While this more traditional approach will not be entirely abandoned (some very important individual contributions will be examined), there are some principled sociological reasons for bringing more attention to the particular nature of the collective endeavours.

In the broader fields of the sociology of knowledge and the sociology of the professions, and certainly in the more refined fields of the social studies of science or the sociology of scientific knowledge, there is a strong recognition that making knowledge visible, credible and ultimately 'legitimate', is a collective endeavour that involves the slow establishment of expertise and the sanctioning of institutionally-backed, methodological

---

\(^{12}\) By 'individual' I do not mean 'only one author'; rather, it refers to contributions to the field that are not intended to speak or represent a collective voice of the field as a whole, but are 'individual' contributions to it. Often collective endeavours of the type I will be reviewing, include individual contributions (sometimes very influential ones) as part of their process and their dissemination strategies. Of course, the legitimacy and credibility of these collective efforts stands or falls on the specific nature of each such enterprise.
protocols and methods aimed at the public performance of 'reliable and valid knowledge claims'. In medical research this has taken the form of various collective processes of 'guideline' establishment and review synthesis protocol development, influenced by the evidence-based medicine (EBM) movement (Eccles & Grimshaw, 2000). More recently, some sociological analyses of these processes have begun to appear (Berg, ter Meule & van den Burg, 2001; Moreira, 2005). In public health, a 'dramaturgical analysis' (Goffman, 1959) has been done of some controversial reports on public health issues by the National Academy of Sciences (Hilgartner, 2000). In the field of health promotion, such efforts at consensus-based methodological protocols are much less developed and institutionalized, and with some exceptions, constitute a 'reaction' to the dominant mode of research synthesizes and guideline development methodologies and processes. In addition, it should be acknowledged that some health promoters feel that engaging in the 'evidence game' may not be worth the candle (Kickbusch, 1997). Without going into a lengthy justification of why I think the latter position is untenable, suffice it to say that this paper will start from the assumption that playing the evidence game is part and parcel of playing the health promotion game and that it is only by radically redefining what it is to 'play' health promotion that one could excuse oneself from the evidence debate.13

---

13 There are two key reasons that a demand to substantiate the effectiveness of health promotion interventions will continue to dog the field: first, no matter how convinced one is of the value of an approach or program, when it comes to 'investments', these activities must compete with alternatives, whether related to health or not, and this requires some form of justification that even a sympathetic policy alliance must produce; second, general theoretical rationales and examples of promising interventions will not suffice for widespread implementation, as the latter requires more detailed evidence concerning why and how particular interventions work or do not work. That governments continue to be concerned about these issues and continue to look to the health promotion community for answers (Hills, O’Neill, Carroll & MacDonald, 2004) should in itself be a signal that ignoring the problems outlined in this chapter is not very helpful; particularly when some of the same voices that advocate side-stepping the issue continue to look to government as the main source of funding for health promotion!
Section 1: Substantive Literature on Health Promotion Effectiveness

The nature of each of these collective efforts ranges across a spectrum, from scholarly activities represented in: 1) an edited book, initiated by a specific group of academics; 2) collaborations with the European Union initiated by the International Union for Health Promotion and Education; 3) collaborations initiated by the European Office of the World Health Organization; 4) an official Task Force on Community Preventive Services set up by the US government; 5) the attempt by the Cochrane Collaboration to integrate findings in the health promotion field to their more general strategy of offering rigorous systematic reviews for informing policy and practice; and finally, 6) The IUHPE’s Global Programme on Health Promotion Effectiveness. Of course, by whom and exactly how, these processes were initiated, is only one dimension to consider in thinking about the different nature of such activities: membership composition and recruitment, organizational structure, relative openness and sharing of information, decision-making processes, and dissemination strategies, are just a few of the multitude of ways one could analyze these collectives.

For the purpose of this review, the limitations of space and methods available (document review)\(^\text{14}\) will allow only a cursory analysis of some of the main outcomes of these various processes, and it will focus more specifically on the implications and impact of these varied efforts on the credibility and legitimacy of health promotion research as a scientific basis for establishing the relative effectiveness of health promotion activities.

*Health Promotion: New Discipline or Multi-Discipline?*

The first collective effort that is reviewed is actually a fairly typical edited collection of academic papers, titled *Health Promotion: New discipline or Multi-Discipline?*

---

\(^\text{14}\) A whole research program could be initiated based on asking how exactly these sorts of processes take place in the field of health promotion.
(Edmondson & Kelleher, 2000). The collection explicitly addresses the impact and influence on the field of a WHO-organized global conference on health promotion, held in Jakarta, in 1997, and sees itself as directly related to policy related questions (of course, this is endemic in health promotion, as many of its most vocal interlocutors are also employed government officials). The book provides a good opening for a variety of reasons, but foremost is the fact that two of the early chapters provide starkly opposing views as to how the field should go about establishing both its own scientific credibility and the substantive effectiveness of its various programmatic activities and interventions. While these two visions are not exactly polar opposites, they do set out many of the challenges and methodological themes that will be considered throughout this chapter.

The first paper, by George Davey Smith, Shah Ebrahim and Matthias Egger, is titled, provocatively, ‘Should Health Promotion be Exempt from Rigorous Evaluation?’ Never mind that the title itself could be accused of some serious question begging\(^\text{15}\), the authors outline a strongly worded argument that health promotion’s claims that it should be evaluated ‘differently’ from the dominant mode of evidence-based medicine, characterised by using the results of randomized controlled trials (RCTs) as the ‘gold standard’ of evidence, are just so much special pleading. Smith et al. rehearse the usual benefits of using randomized, controlled designs, versus non-randomized, observational studies, particularly with the focus on eliminating selection bias and confounding, which they later ‘demonstrate’\(^\text{16}\) in relation to a series of results that health promoters have regularly claimed is evidence of the

\(^{15}\) It is precisely the question of what should count as ‘rigorous evaluation’ that is at issue; assuming that the RCT is the only rigorous method of evaluating health promotion is question begging.

\(^{16}\) Despite the call for ‘rigour’, there are several unsupported leaps in logic made by the authors to reach their conclusions.
effectiveness of particular health promotion approaches, such as multiple risk factor interventions.

This is the real challenge to health promotion from the defenders of so-called ‘rigorous’ evaluation: you can show some change here, but how do we know that this change didn’t happen because of confounding factors you were unable to control for? Thus for example, Smith et al. argue that the large claims made in favour of the North Karelia program, inducing reductions in cardio-vascular disease (CVD) risk factors and mortality rates, are in ‘fact’ (‘appear to be’) due to the more general secular trends in mortality reduction across Finland over the same time period. It is important to carefully consider just what is and is not being claimed by Smith et al. It is equally important to highlight the rhetorical slippage that is ubiquitous in ‘scientific evidence debates’ such as this one. First, all they have shown is that a secular trend is a plausible alternative explanation for declines in mortality to the attribution of the effect to the health promotion interventions in North Karelia. While this is certainly legitimate, they do not address the fact that community-wide, multiple strategy health promotion programs are almost inevitably contagious and ‘leaky’; that is, using reference populations can be problematic because you cannot assume that the reference population was not ‘contaminated’ by the intervention (Nutbeam, Smith, Murphy & Catford, 1993). Thus, a rigorous ‘control’ group is rarely feasible without detrimentally affecting this type of community-wide intervention itself. Conversely, there exists evidence to support the idea that the North Karelia experiment did have wider effects on the whole of Finland (Puska, Tuomilehto, Nissenin & Vartiainen, 1995).

They do not and cannot claim that they have direct evidence that the multiple risk factor interventions do not work. This is because as they admit:
"This evidence [their study] does not prove the null hypothesis (i.e. it is not possible to say that there is no benefit from multiple risk factor intervention) but, rather, the evidence does not refute the null hypothesis and does not give support to the alternative hypothesis of attributing benefit to the intervention" (p.23).

Yet, technically, the ‘null hypothesis’ can’t have been about whether or not there is benefit from multiple risk factor interventions. The null hypothesis must have been that there were differences in benefit from being involved in the intervention and being involved in the control group (and by definition, not receiving the intervention). The ‘significance test’ is then to see what the probability is that the observed differences are due to chance and thus are consistent with the null hypothesis. If this probability is small (conventionally, less than 5%), then it is not consistent with the null hypothesis and supports the alternative hypothesis that the differences are not due to chance and (depending on the robustness of the design) are due to the intervention. This means that the authors have not been able to support the alternative hypothesis. In other words, the interventions they reviewed may very well have had an effect, but according to this particular review, the differences observed could have been due to chance, or at least did not meet the conventional standard of significance that would make chance causing the differences a very unlikely scenario. This peculiar way of assessing evidence (it is, of course, the standard way) is not immediately intuitive; that is, it is not clear to the uninitiated why exactly this way of going about deciding how strong evidence is should be the best way. By simply making this ‘strange’ we can bracket off normative judgment for long enough to reconsider the assumptions that have led to such a ubiquitous orthodoxy in the medical and health fields.

The first thing to notice is that if one digs into the entrails of any complex example of meta-analysis as the one referred to in this article (Ebrahim & Smith, 1997), one immediately realizes that what is reported as quite a straightforward finding ("The effect of multiple risk
factor intervention on both total and CHD mortality was insignificant”) is anything but. There are a variety of complex statistical techniques (pooled estimates, odds ratios, random effects analyses) based on a thick network of assumptions, that allow the data to be transformed into analytical results, which only then can be interpreted to arrive at the simple conclusion. All these steps and assumptions are contingently valid\textsuperscript{17}, depending on the situated reasonableness of the arguments the authors put forward.

All these technical pyrotechnics (and like real pyrotechnics, they are prohibitively expensive and become more so the higher the quality), are aimed at only one particular epistemological difficulty: how do we know that an intervention has produced beneficial effects? Even narrower, it is really concerned about the worry that we might make such attributions, when actually, other factors we haven’t accounted for really produced the benefit. As will be discussed, this epistemological worry is based on a particular philosophical ontology that is not the only option. Whatever the case, the appeal of this push for ‘rigour’ is powerful and can quickly lock people into the particular frame of reference that identifies ‘rigour’ with the RCT methodological solution. So where does this leave our ‘evidence’ problem? It must be acknowledged that, on their own terms, Smith et al. present a major challenge for health promotion evaluation; yet, it is only a very partial story that ironically incorporates a particular ‘selective bias’ that privileges one epistemic issue above all others, and completely ignores any of the cogent ontological and epistemological questions raised by the RCTs critics. Nevertheless, the paper brings some sharp focus to the debate and is uncompromising in its dismissal of the alternative routes some others advocate

\textsuperscript{17} Not the techniques themselves, which are ‘statistically valid’, but the application of them to this particular problem and the type of data available for analysis. The contingency arises because ‘data’ do not determine analytic technique.
for examining the effectiveness of health promotion, specifically it holds its ground with
most of its critics in its claim to be the only way to establish a sound causal claim that an
intervention has had specific effects.\(^1\)

The second article in this collection that will be considered is the one by Keith Tones
(2000). Tones begins by acknowledging that health promotion should be subject to rigorous
scrutiny as to its effectiveness. However, he rejects the approach and methods demanded by
Smith et al. as “largely irrelevant since they fail to acknowledge the peculiar characteristics
and complexities of health promotion” (p.29). Tones sets out to challenge the relevance of
the RCT on the bases of three related points:

- “the fact that health promotion programmes are inherently more sophisticated and
  complex than the interventions typically associated with clinical trials”
- “the centrality of empowerment and participation in health promotion and the associated
  importance of action research”
- “the quest for illumination which is central to the development of effective programs”

I address each of these points in turn, as they exemplify important themes that reoccur
in many of the contemporaneous and subsequent efforts to tackle health promotion
effectiveness.

**Complexity**

The problem of ‘complexity’ enters at many different levels in the attempt to evaluate
the effectiveness of health promotion. There is much confusion surrounding this key concept.
Some understand it to mean that health promotion interventions are ‘simply’ more
complicated and involve too many different uncontrollable variables to allow for the type of
procedure proscribed by a RCT (Nutbeam, 1999). Others argue against this and suggest that
the problem lies in our not evaluating the right phenomena; if we pay attention to ‘functional’

\(^{1}\) The realist approach is the only one that takes this argument on directly. This will be discussed in section
two.
processes, rather than 'formal' identities, we can use the RCT design to look at complex interventions (Hawe, Sheill & Riley, 2004). However, there are two other distinct ways of approaching the problem of complexity. First, one could truly follow the lead of complexity science and note that the real problem is not the magnitude of potential variables that are important, but the interactions between them. In other words, what matters is that we are using very simplistic analytical devices (the RCT being one) in order to understand emergent phenomena that operate through qualitative changes generated by non-linear, dynamic processes (Waldrop, 1992; Gribbin, 2004), where specific networked configurations and thresholds hold the key to illumination. This perspective has hardly been taken up by health promoters seriously, although there have been some gestures toward such a shift (McQueen, 2001; Dooris, Poland, Kolbe, De Leeuw, McCall & Wharf-Higgins, 2007). Alternatively, complexity can be understood as the problem generated by the relative immaturity of conceptual development within the field. Not being alone in this, health promotion suffers from the generic situation pertaining within the sciences of social life, where so-called 'variables' proliferate, without much, if any, agreement on the foundations of key concepts and measurement devices. On this basis, 'complexity' is related to the dearth of available operationalized concepts and tools that have attained a consensus within the field, in order to carry on reliable and valid measurements, and thus guarantee the substance of established 'data' upon which the RCT method must rely. As we will see, how one approaches this issue, 

---

19 I have yet to understand how this particular solution is helpful (or if it can sensibly be called a solution at all). The RCT (despite the critique developed here) is a wonderfully robust research design, the virtue of which relies precisely on its ability to simplify an evaluation question into a strictly linear causal hypothesis; diluting this elegance through misguided attempts to make RCTs more 'flexible' seems to me not to be a worthwhile project.
defines the types of solutions that one pursues in the search of a more viable approach to evaluating health promotion effectiveness.

*Empowerment and Participation*

There are also a variety of different, although complementary rationales for why the values and principles based on empowerment and participation make it difficult, if not impossible to pursue a RCT-type research strategy for health promotion evaluation. These range across three dimensions: ethical; practical/feasible; and, epistemological. Tones, along with many others, emphasizes the ethical and practical dimensions. Thus, the primary objections to RCTs for health promotion relates to the significant ethical problems with asking communities to participate in research that may not directly result in benefits for that particular community. Using ‘control’ communities becomes ethically problematic to the extent that many communities are not convinced that participating in a RCT will result in beneficial knowledge for their community. In addition, RCTs are constitutionally biased toward summative as opposed to formative evaluation, which, as Tones notes, denies the ability for participants to take advantage of ongoing learning and change their actions in the midst of the evaluation process. There are also strong ethical arguments against RCTs based on the fact that the methods of quasi-experimentation tend to exclude or deny the self-determination of subjects.

Practically, RCTs are very expensive, and in order to control for many variables, the size of the population sample must be large, leading to the critique that extensive and multiple RCTs in the area of health promotion intervention research just isn’t a feasible option. For example, the methodological critique of many contributions to the Cochrane Collaboration in relation to the much more established clinical trials, is that not nearly
enough well designed studies are available. In many areas of health promotion, in order to reach the level of even the least validated of clinical topics, the funds required would far exceed what is now allocated, or even optimistically what could possibly be allocated for health promotion research. Rightly, many communities just will not accept the proportion of funds that advocates of RCTs argue must be diverted to ensuring so-called ‘rigorous evaluation’. As long as community participation and empowerment are core values, it is difficult to see how to generalizing an evaluation strategy as outlined by Davey Smith et al. would be practical. As Tones (2000) points out, even a relatively simple health education intervention will tend to have multiple layers, multiple strategies, multiple staged outcomes and many other complex factors that make well-structured RCTs in health promotion technically and financially unfeasible. Finally, there are serious epistemological questions (not often articulated by critiques of RCTs) concerning the lack of a participatory approach. This revolves around the simple issue Tones raises, that an RCT: “rarely, however, yields data that which provide insight into why particular parts of a programme have been successful or unsuccessful” (p. 40). This relates to participation and empowerment directly because the only way of getting these insights is by understanding from the point of view of the participants how and why certain elements of a program work or do not work. RCTs work to systematically exclude this type of information, based as it is on potentially biased sources. This last point leads to the concern about the purpose of an effectiveness study: what are we trying to find out

_Illumination vs. Summation_

A major point made by Tones, and one that is often repeated throughout the evidence debate in health promotion, is that summative evaluation, focused simply on whether an
intervention achieved pre-determined outcomes, is not the most important information to find out about most health promotion interventions. It tends to obscure many key questions such as: whether an intervention has been properly implemented; what parts of a program or project were the key elements in achieving success or inducing failure; which intermediate outcomes are the most relevant for assessing effectiveness; what factors provide the basis for the sustainability of the program. With RCT it is difficult to address these questions, and the more complex one makes the RCT, the more expensive and the less easy to interpret the results. This latter issue has profound implications for the synthesis of evidence, as the main statistical techniques for meta-analysis are much easier to use with basic, homogeneous outcome measures. Raising the question of ‘illumination’ as a research objective means to challenge our thinking about effectiveness and asks us to consider alternative ways of assessing whether or not health promotion has achieved or can achieve its goals. For example, it is often mentioned that health promotion is meant to be about empowering people to have control over the determinants of their own health, yet rarely do we evaluate health promotion activities on that basis. Furthermore, if we did so, it is very unlikely that the usual methods of assessment would be appropriate.

Between these two articles a profound chasm exists concerning how we should go about assessing and evaluating health promotion effectiveness. What is clear about the book in which these articles sit, is that it is not untypical of the normal academic collection, in that there are opposing views presented, with no particular attempt to arrive at any consensus about the results of their inquiries, nor about what approaches are recommended for pursuing further research. This is not true for the remaining collective efforts to be briefly analysed. Each of the latter, are efforts to forge some level of consensus about these questions and to
present recommendations for future progress in the field. All of them, in one way or another address the problems that face health promotion as starkly outlined in the disagreement between Smith et al. and Tones. In fact, it is precisely the fact that such large disagreements can exist that creates a credibility and legitimacy problem for health promotion researchers and leads to significant efforts to come to a more common and authoritative understanding of the appropriate way forward.

*IUHPE: The Evidence of Health Promotion Effectiveness*

I start the review of these consensus-oriented efforts with an initiative of the International Union for Health Promotion and Education. It is notable as a directly political effort to persuade politicians and policy-makers that important evidence did exist in relation to the effectiveness of health promotion after 20 years of program implementation. I will focus here on the culmination of the project, which resulted in a report to the European Commission (IUHPE, 1999). Although it does address, in its first chapter, methodological issues, it is primarily focused on case studies and recommendations. For the purpose of this paper, the first chapter will hold disproportionate attention and most other comments are of a generic sort summarizing the overall nature of the project.

The project had a Project Advisory Group, made up of top international health promotion experts, a group of Lead Project Authors responsible for the case studies, and a ‘Witness Group’ of political experts from across Europe. In its own terms, and according to its immediate objectives, the project was very successful, in that it produced a well-organized and succinct document that was clearly presented and disseminated for its primary target audience. However, from a methodological standpoint, there are some significant issues that were either not addressed or were sidestepped. To begin, we can examine Don Nutbeam’s
opening chapter, which is clearly meant to set the stage for how to understand the issues and challenges of assessing health promotion effectiveness. Nutbeam addresses four key questions:

- What do we mean by effectiveness in health promotion?
- What measures should be used to assess effectiveness?
- What research methods should be used to evaluate effectiveness?
- What programmes are worth evaluating?

Nutbeam argues that all these questions are related to the entry point (issues, populations or settings) of a health promotion intervention, its method of implementation (individual-level, single strategy, small scale to multi-level, multi strategy, large scale and combinations of these) and the intended audience (policy-maker, practitioner, academic, etc.). He develops an argument that health promotion should look at different stages and levels of outcomes and should use methods appropriately tailored to specific types of interventions. Many of Nutbeam’s arguments and recommendations have been taken up by others and form the background to much of the consensus-building work that we shall see later; however, what is not clear from this chapter, nor from the subsequent case studies, is what alternative approach to synthesizing evidence is being used or advocated for. Each chapter stands on its own in terms of how it arrives at the analytical conclusions it reaches and what level of quality it attributes to the particular evidence it reviews. From the standpoint of a Davey Smith one would see much to criticize in this procedure. There is a sense in which the type of syntheses represented in this report suffers from the particular transitional nature of the debate at the time. In other words, the appeal to more flexibility in the type of evidence and methods to be used (chapter 1) is married to a set (the individual case studies) of uncertain and inconsistent approaches to synthesizing evidence. What we are left with are some fairly persuasive narratives that stitch together a variety of different
evidential sources, but which because they do not break explicitly (and offer an alternative justification or methodology) with the dominant form of evidence synthesis, end up being open to a strong methodological critique that compares such efforts unfavourably to the more 'rigorous', systematic review, especially the meta-analytic approaches. Thus for example, typical review 'tables' are constructed (with study type, subjects, intervention and outcomes) which only serve to highlight the inadequacy of the evidence in comparison with pooled estimates and odds ratios that one finds in meta-analytic reviews (IUHPE, 1999, pp. 23-25).

It should be pointed out, that on this playing field, the systematic reviewers of the EBM movement are always going to make such arrangements of the 'evidence' look paltry. The methodological arguments for avoiding the non-systematic approach to reviewing evidence are unassailable on their own terms (Chalmers & Altman, 1995; Jackson & Walters, 2005; Mulrow, 1987; Stevens & Abrams, 2001). Compiling favourable outcomes and plotting them in a table looks to the systematic reviewer a textbook example for why their services are necessary.

None of this is to disparage the work of the report; it is more to point to the difficulty facing health promotion when playing the evidence game. It is not the 'evidence' that is missing, but the procedures by which the evidence is selected and the conclusions arrived at.

Ironically, what this work did provide was not so much a solid basis for the scientific credibility of health promotion, but rather provided a platform for further work on the effectiveness of health promotion, particularly through the stimulus of the follow up to this project in IUHPE's Global Programme on Health Promotion Effectiveness (GPHPE), which will be the final collective effort that will be analysed here. Next to be considered will be the
twin effort, led by the WHO European Office on evaluation in health promotion that reported almost contemporaneously.

**WHO: Evaluation in Health Promotion**

In essence, the book Evaluation in Health Promotion: Principles and Perspectives (WHO, 2001), known as “the yellow book”, was an attempt to fill the gap that is only signalled in Nutbeam’s opening chapter of the Effectiveness book. That is, it was specifically oriented to “stimulate and support innovative approaches to the evaluation and practice of health promotion, by reviewing the theory and best practice of evaluation [my emphasis]” (3). The Regional Office set up a Working Group on Health Promotion Evaluation with the support of the Centers for Disease Control (US), the Health Education Authority (UK) and Health Canada. The Group had four products, one of which was the “yellow book” which was developed on the basis of over 30 commissioned papers (not all of which were used in the final document). Significantly, the Group decided to establish ‘basic principles’ of health promotion that would guide their work and their recommendations. The Group listed the principles that characterize health promotion activities as:

- empowering
- participatory
- holistic
- intersectoral
- equitable
- sustainable
- multi-strategy

In addition, there were four ‘core features’ identified: participation; multi-disciplinarity; capacity-building; and, appropriateness. Using these guiding resources, the Group managed to pull together many important papers, now recognized as seminal in the field. They exceeded their goal of stimulating innovative approaches and the field is only
beginning to follow up on some of the more suggestive leads provided by many of the 
authors. Furthermore, although the arguments presented in each chapter are heterogeneous 
and not entirely consistent with each other, they did provide enough consensus around certain 
issues that the Group could confidently assert some recommendations and offer a compelling 
synthesis concerning what direction health promotion evaluation should be headed 
(Goodstadt, Hyndman, McQueen, Potvin, Rootman & Springett, 2001). The most important 
accomplishment of this collective endeavour was to give much needed credibility to what 
were often marginalized efforts to produce appropriate alternative evaluations of health 
promotion activities. It tackled directly the ontological, epistemological and methodological 
problems and issues at the heart of the evaluation debate and provided a strong argument that 
RCTs were not the appropriate method to evaluate most, if not all health promotion 
interventions. A fortiori, it undermined the idea that RCT should provide the ‘gold standard’ 
of evidence in the field. Thus, it managed to ‘open up’ the restrictive nature of the evidence 
debate as dominated by the EBM movement, and in a sense left us with the injunction to “let 
a thousand flowers bloom”, so to speak.

Nevertheless, there are two significant (and acknowledged) problems that remained in 
relation to the problem of knowledge synthesis. First, although it was advocated that 
‘qualitative’ approaches should be used more readily as the appropriate ones for many health 
promotion interventions and questions, it is not clear just what is meant by qualitative: is it a 
different ‘paradigm’, a different ‘methodology’, or just a variety of different ‘methods’. This 
makes it difficult to be more precise about recommendations and thus makes the future of the 
‘evidence-base’ being created just as diverse and unwieldy as it is now, even if it is more 
populated with evaluations. This leads to the basic issue concerning the synthesis of this
evidence. For, even if these various approaches gain credibility and produce findings, how are we to combine them into broader statements concerning the effectiveness of health promotion approaches, using credible methodologies and methods of synthesis? To give the Group credit, they never tied to avoid this problem, it was just that their focus was mainly on the development of better approaches to producing primary evaluation evidence and did not specifically alight on review syntheses as a sui generis methodological dilemma. These questions are analogous to the perennial challenges within various so-called ‘qualitative’ approaches to address the issues of reliability, validity and inferences from single cases to the general case. It is not that potential solutions are totally unavailable; rather, it is that making inferences across extremely heterogeneous evaluation data demands more methodological rigour, not less, than the meta-analytic technique. What the ‘yellow book’ produced was a map of the unknown, where broad areas of knowledge development are pointed to, but the detailed cartography is left for future explorations. As noted in a key chapter, the health promotion field has yet to resolve what is to constitute ‘evidence’ and how to rank its quality, nor has it attained a level of theoretical maturity that would underpin such a resolution (McQueen & Anderson, 2001). The implications of this immaturity can be seen in the next two cases, where attempts to develop review syntheses of health promotion interventions have had to retain a methodological infrastructure that is incongruent with the type of phenomena it aims to investigate and incongruent with the type of ‘evidence’ likely to be developed if the recommendations of the ‘yellow book’ were faithfully carried out.

*The Task Force on Community Preventive Services: The Community Guide*

The Task Force on Community Preventive Services, was set up by the US Department of Health and Human Services in 1996 and is by far the longest running and
most well funded of all the research on the effectiveness of health promotion. The Task Force is an independent body, but is intimately linked with the CDC and derives much of its resources and scientific advice from that agency. The Task Force’s first job was to develop and refine a “standard methodology to search for and sift through evidence from a wide range of studies reflecting many disciplinary perspectives” (Zasa, Briss & Harris, 2005, p.xii). This quote reflects two difficult but inevitable aspects of their work: one, a project dedicated to systematic reviews of existing evidence could not help but establish clear methodological guidelines to retain rigour and thus credibility; second, it had to work with an extremely diverse range of materials, most of which may not rise to the standards that the methodological protocols would set. What is probably most interesting about this massive endeavour is not that it utilised a fairly standard methodological protocol for conducting its reviews (given its mandate it would have been difficult to do otherwise), it is that it faced up squarely to the inevitable limitations imposed by this choice. The Guide the Task Force produced is a model of transparent and self-critical scientific work. A superb concluding part to the book on ‘Methodological Background’ outlines in detail the step-by-step processes of the reviews and the rationale for each step and ends with a very interesting chapter on ‘Continuing Research Needs’. The Guide adopts the standard evidence hierarchy, even though it acknowledges that often RCT evidence is unavailable for population level public health interventions. This means that many of its reviews are characterised by both an open attitude to the type of evidence it will consider, followed by the inevitable winnowing out of studies due to ‘threats to validity’. This means that many areas examined are characterised by the finding of ‘insufficient evidence’ to make a recommendation. For example, in the crucial area of ‘The Social Environment’, 6 out of 8 interventions examined are said to have
insufficient evidence to make a recommendation. Furthermore, there seems to be an inverse relationship between the complexity of the intervention and the amount of acceptable studies for inclusion in the review. There is an irony here that should not be lost on the health promotion community. It is precisely these more complex, upstream interventions that are consistently advocated for as consistent with a health promotion approach. It seems the closer we get to the intent of the Ottawa Charter vision, the farther we are from meeting the inclusion criteria in the world of systematic reviews. There are a few points to note about what underlies this approach to systematic review: 1) the overarching emphasis is on study design, not on the details of the case itself; 2) the unit of concern is the study as a whole, not any particular knowledge claim within or across studies; 3) the definitions of validity that are appealed to are derived from a specific orientation to science, not one shared by all evaluation approaches. The quality criteria are heavily biased towards a quantitative approach, and weaknesses are identified in relation to criteria that are hopelessly inappropriate if applied to evaluative evidence that is qualitative in nature. The suitability of study design is determined mainly in relation to threats to internal validity, meaning the "confidence that the intervention being evaluated really caused the effects or outcomes being measured" [my emphasis] (p.438). Although external validity is also assessed, there is a priority given this causal question. As this notion of causality is a central methodological concern, it is worth dwelling on it once again to mark this topic for the later more intensive methodological discussions. There is an extraordinary emphasis on concerns regarding an epidemiological approach to evidence of causal effect, particularly given that the problem with most complex health promotion interventions is that causal relationships are obscured rather than illuminated by the naming of a type of intervention; most often, the theoretical
causal chain is not well specified, so it is difficult to tell just what ‘exposure’\textsuperscript{20} is being evaluated, never mind what level of exposure.

A thought experiment might help clarify this conundrum: imagine that we have one, impeccably designed RCT study, examining the implementation of ‘an intervention’; from this study we are asked to interpret the likelihood of a true causal relationship between the ‘intervention’ and some putative outcome. Unfortunately, we have no idea, actually, just what ingredients of the intervention are meant to be key in making change happen. On the other hand, we have a ‘poorly’ designed study (i.e. one without a control group), but one which outlines in detail the precise mechanisms said to be active in producing the observed change. From an epidemiological point of view, we have the problem of possible biases and confounding; however, from the broader point of view of causal analysis, we have much more relevant information on which to base our judgments about effectiveness. We then have a basis on which to make a causal claim, even though we may still want to do further studies, including appropriate epidemiological investigations, to increase our level of confidence in the suggested relationship. This is, of course, merely a thought experiment, and there is no particular evidence that many such alternative non-comparative studies exist that would be as detailed and as persuasive as would be necessary to convince us of a causal link; nevertheless, the point is that even if a plethora of this type of evidence did exist, it would be ruled out of court by a methodological fiat determined by a fixation on an epidemiological interpretation of the issue of causality. What the Guide demonstrates is that even a process that is open, self-critical and fair in its processes and its analytical judgments can

\textsuperscript{20} It is not even clear that this type of language, derived directly from epidemiological/bio-medical ontologies, is appropriate to the type of causal relationships that might pertain in many health promotion processes. For example, if participatory processes have any causal effect, they are not the type of phenomena that ‘happen to’ people in the sense of an ‘exposure’.
systematically bias its approach through an unacknowledged attachment to assumptions that are not themselves defended. While one of the Guides' great accomplishments is to point out some major gaps in the evidence base of health promotion, it's particular epidemiological bias in methodological strategy promises to continue to pass over potential sources of evidence in the future.

The Cochrane Collaboration: Systematic Reviews of Health Promotion

The Cochrane Collaboration is the very epicentre of the EBM movement, wedded inextricably to the evidence hierarchy and the main driving force behind imposing 'rigorous evaluation' on all health interventions as defined within the context of the dreaded 'gold standard' of RCT evidence. Yet, the story of the Health Promotion and Public Health Field (HPPH) of Cochrane is an interesting one. We see here the first signs of a significant shift by one of the bastions of EBM towards considering some alternative strategies for synthesizing evidence. This can be attributed to a variety of factors, amongst which are the positive efforts represented by the 'yellow book', the negative examples in terms of results represented by the Community Guide, and the actual combination of particular individuals that have been involved in this specific Cochrane process. It is clear that the HPPH section of Cochrane have made a concerted effort to bring many of its severest critics inside the tent in order to help it address some of the well-articulated concerns regarding evidence in relation to health promotion. The Cochrane Qualitative Research Methods Group (CQRMG) has also been making headway in addressing the issue of including research studies in review syntheses and developing guidelines for assessing the quality of these studies. While the HPPH has not (unsurprisingly) entirely abandoned the epidemiological and positivist bias in its evaluation of what should count as evidence and how to go about assessing effectiveness, it has moved
towards addressing many of the important critiques of the EBM, evidence hierarchy model. The HPPH group has recently produced two key documents: Guidelines for Systematic reviews in health promotion and public health interventions (Jackson & Waters, 2005); and, a corresponding Handbook (Jackson, 2005). The group is based out of and partially funded by the Victoria Health Promotion Foundation (VicHealth) in Melbourne, Australia and is also supported by the Department of Health, UK. Its academic institutional support comes from Deakin University (Aus) and the Evidence for Policy and Practice Information and Co-ordinating (EPPI) Centre at the University of London (UK). The Guidelines were produced by an international Task Force of leading experts in the field of health promotion and public health. Although the Task Force did not manage to produce a comprehensive set of guidelines for synthesizing evidence generated from a variety of different methods, they were able to point to emerging attempts to do so and they were able to map out many of the methodological issues to be resolved and key developments that are needed. Perhaps most importantly, there is a noticeable weakening of the normal evidence hierarchy. For example, the Guidelines follow Petticrew and Robert’s (2003) suggestion that the review question should determine what type of study design is most appropriate and Glasziou, Vandenbroucke and Chalmers’ (2004) that “criteria used to select studies should primarily reflect the question/s being answered in the review, rather than any predetermined hierarchy” (Jackson & Waters, 2005, p.6). The Guidelines argue that non-randomised controlled trials may offer the best evidence at the time of a review, but more interestingly there is a strong advocacy for the use of qualitative research in effectiveness reviews and offer some suggestions as to how this research can be integrated into a systematic review. This goes a long way towards bringing credibility to the type of evidence advocated for in the ‘yellow
book’, and follows up other efforts that have recently been accumulating in the literature (Glasziou et al., 2004; Murphy & Dingwall, 2001; Popay, Rogers & Williams, 1998).

However, despite this significant shift in attitude, there are still some major questions that are left unanswered. For instance, the ‘evidence hierarchy’, although not enforced (mainly for pragmatic reasons), still lingers in the background and thus no major reconsideration of methodology is suggested and certainly nothing like a ‘paradigm’ change argued for. Like most of the ad hoc efforts to ‘mix methods’ there is much exhortation to be pragmatic, but little rigorous guidance on how theoretically divergent approaches are meant to be stitched together to make up a satisfactory whole and answer specific research questions. This very admirable work strikes one as a near perfect analogy and actually sensible use of part of Kuhn’s work on ‘scientific revolutions’ (1970). The reaction to the accumulation of anomalies presented by trying to fit health promotion into the EBM paradigm, is exactly as expected from Kuhn’s analysis of the ‘response to crisis’:

“.....first what scientists never do when confronted by even severe and prolonged anomalies. Though they may begin to lose faith and then to consider alternatives, they do not renounce the paradigm that has led them into crisis. They do not, that is, treat anomalies as counter-instances, though in the vocabulary of philosophy of science that is what they are” (p.77).

We see very clearly in the HPPH group an attempt to deal with clear anomalies presented by the health promotion field; yet, these ‘problems’ never shake at the very heart of the evidence-based movement’s foundation: the paradigm is as yet, secure.

*The Global Programme for Health Promotion Effectiveness*

The GPHPE was initiated by IUHPE as part of a follow up to the original Effectiveness project. It was felt that the first project, although successful in its own terms, actually raised many questions of a substantive and methodological nature. A major weakness of the first project, as well as of the ‘yellow book’, was their ‘Western bias’; that
is, all of the evidence and the methodological work considered came from Europe, North America and Australia. Second, while some traction was gained by playing the evidence game along the lines of the EBM perspective, it was clear that more work needed to be done to develop alternative approaches to evaluating effectiveness and synthesizing evidence. Third, there was quite simply a paucity of evidence in many areas, whatever ones preferred definition. In order to facilitate work that addressed these issues, IUHPE, under the direction of David McQueen, decided to organize itself along the global IUHPE regions and let each region produce its own contribution to the overall work. There was very little central control or direction of the nature of each regional project and so very divergent sets of activities ensued. The latter strategy was explicit and was meant to address both the Western bias in terms of pure coverage and to avoid imposing methodological restrictions that would preclude any important innovation. The results of this endeavour are slowly coming to fruition and recently produced a monograph entitled, Global Perspectives on Health Promotion Effectiveness (McQueen & Jones, 2007). However, there are a few preliminary comments we can make about this programme. First, the idea behind it is actually quite radical, in that by its very nature, the results are developing ‘organically’ and not by any predetermined design. This means that some regions have been able to grow their project much more quickly than others; furthermore, some regions have found it necessary to focus on ‘preparing’ the ground for any effectiveness research activities at all. Second, even in those regions that have attempted to further alternative methodological approaches (Europe and North America), the progress has been difficult and what has been discovered is that much conceptual ground work remains to be done before proclaiming any new panacea for assessing effectiveness.
As McQueen himself concludes (2007), the phenomena of complexity, contextualism, and reflexivity all seem to be at the heart of the real-world life of multi-layered, multi-actor, multi-strategy health promotion programs; however, all these ‘truths’ only make the evidence and effectiveness ‘game’ harder to play. Along with other recent efforts to further develop theory in health promotion (McQueen & Kickbusch, 2007), we see a convergence of opinion around the challenges health promotion faces in the evaluation of effectiveness realm; yet the possible solutions are only now starting to be mapped out and the Global Programme is another major initiative that has allowed the space for this halting progress to continue.

The huge benefit of this large, unwieldy and meagrely funded programme has simply been to give credibility to research explorations that would otherwise have found it very difficult to avoid being dismissed out of hand. This has meant that much important thinking has been generated about the depth of the methodological troubles generated by the evidence question in health promotion, but also about some of the potential pathways to be followed in the future.

Summary

Overall, we see from these six examples a variety of collective efforts to address the issue of demonstrating the effectiveness of health promotion programs and interventions. There is an obvious struggle with a prevailing approach to establishing what evidence is, and how to synthesize that evidence. The temptation has been to establish credibility through adherence to the established methodological protocols of EBM, rather than through fidelity to the nature of the phenomena under consideration; however, a shift can be detected, even at the core of the evidence-based world. The felt need to cling to the ‘paradigm’ of EBM and its evidence hierarchy is more than understandable, considering that no well-established and
credible alternatives are on the horizon. This means that in order to better understand the potential for developing such an alternative, one needs to more deeply consider the basic philosophical assumptions underlying the dominant approach and to face squarely the challenges that an alternative would need to meet if it was to attain even a modicum of EBM's credibility.

Section 2: Methodological Troubles

Introduction

What follows is an analysis of three key orientations/‘paradigms’ that are relevant to the evidence debate in health promotion:

*Positivism, Epidemiology and Evidence-based Medicine*: the evidence-based medicine movement, epidemiology and positivist philosophy of science.

*Constructivism, Phenomenology and Health Promotion’s Reaction*: the health promotion reaction and the naturalistic/constructivist ‘paradigm’ of evaluation.

*‘Post-positivism’, Critical Realism and Realist Synthesis*: the search for a ‘middle ground’ and the critical realist alternative.

The scare quotes around ‘paradigm’ are meant to signify a strong element of doubt about the use of the term employed by the influential work of Guba and Lincoln to characterize different ‘belief systems’ in the social sciences (Guba & Lincoln, 1989; Lincoln & Guba, 2000). There are some very good reasons to restrict the use of this term to the one Thomas Kuhn gives in his *The Structure of Scientific Revolutions* (1970), where a paradigm is intimately tied to the idea of ‘normal science’, where basic perceptual categories and
concepts, not just ‘belief systems’\textsuperscript{21}, determine how scientific practitioners go about there normal work. Paradigms, in this sense, are not things one can simply change by changing beliefs, but are usually only revolutionized by slow generational change; in fact, often the key generators of specific anomalies in previous paradigms are unaware of their future impact.

The only orientation that approaches a ‘paradigm’ in the Kuhnian sense is the first one, organized around positivism, epidemiology and evidence-based medicine. The key here is that something approaching a coherent set of epistemological positions, theories, and methodological practices that focus on ‘puzzle-solving’ activities, aimed at producing findings and results that are recognized across the community of scientists involved in the work. It is these ‘practices’ that become second nature, to the extent that it just simply \textit{is} the way you do ‘scientific evaluation’. In relation to the issue of assessing the effectiveness of health interventions, a clearly shared agenda and set of workable problem areas have been identified and ‘solutions’ are regularly arrived at and recognized as such. These conditions, to put it mildly, are not satisfied for either the mélange of approaches that are often lumped under the notion of a ‘constructivist paradigm’, or for the critical realist alternative. In fact, one of the arguments of this thesis is that it is precisely the ‘pre-paradigmatic’ character of the alternatives that makes them difficult to prop up in opposition to EBM.

Each of the three sections is structured into two parts:

- An introductory paragraph, providing an overview of some of the main critical issues associated with each broad orientation. This is followed by an analysis of these orientations, along four dimensions: i) \textit{Ontology}; ii) \textit{Epistemology}; iii) \textit{Methodology}; and, iv) \textit{Ethics}.

\textsuperscript{21} It is important to understand that, for Kuhn, a \textit{paradigm} rests on a basic phenomenologically shared perspective. ‘Beliefs’ are second order, reflective concepts.
• Some implications of these various philosophical and methodological commitments for developing synthesizes of evidence in health promotion.

*Positivism, Epidemiology and Evidence-based Medicine*

Often it is said that ‘positivism’ slavishly adopts the methodology and methods of the natural sciences for studying human phenomena. It then is necessary to demonstrate how different the phenomena of the natural and human sciences are in order to show the wrong-headedness of positivism. In a somewhat contrarian manner, this works starts from the premise (evidenced in much of the recent work in the history and sociology of science) that the intensely methodological concerns of positivism, epidemiology and evidence-based medicine are actually unusual for most of the mature sciences, both in their origins and in their contemporary development. The philosophical self-understanding of natural science has always been the cloistered preserve of philosophers of science and has little troubled either the substantial theoretical progress or the refinement of method within physics, chemistry or biology.

Positivism as a philosophical approach, epidemiology as a emerging science, and evidence-based medicine as a sort of high church, methodological policing movement, all derived, not from the natural sciences, but from a deep anxiety about the epistemological status of correct inference in less established disciplines (Porter, 1995). This anxiety is rooted in a very particular philosophical history, with a peculiar ontology, epistemology, methodology and even ethics.

i) The *ontological* position that underlies the positivist philosophical program is not, as Guba and Lincoln incorrectly state, a ‘realist’ one (1989). Quite the contrary, positivism emerged from a specific outgrowth of skeptical empiricist thought, based mainly on the assumption of an anti-metaphysical ‘irrealism’. For consistent empiricist-positivists, any talk
of substantive entities existing outside of human experience was simply out of bounds for serious philosophical reflection; the only ‘evidence’ we have is that which is available to our senses, so to speak of ‘independence’ or ‘single realities’ was nonsensical, metaphysical speculation and something to be avoided at all costs.

ii) The epistemological position that derived from this severe ontology meant that the only hope for objective knowledge was to establish clear, unambiguous criteria for intersubjective inquiry. It is a complete misunderstanding of positivism and its empiricist heritage to think that its objectivism derives from a naïve assumption about a mind-independent reality; such a travesty of what was at stake for the positivists misses the deeply reflective and analytical work of amongst the most brilliant minds of their generation\(^{22}\). Even Popper, who distanced himself from the Vienna Circle, worked explicitly from the empiricist-positivist problematic; falsification, based as it was on the empirical testability of theoretical hypotheses, was clear that ‘established fact’ (which was the empirical data against which to test the theories) was an inter-subjectively constructed phenomena, derived from agreed upon methodological protocols amongst a community of inquirers. This contrasts to the usual construal and simplification of positivism as ‘subject-object’ dualism\(^{23}\). Thus, positivists were, ironically, the first ‘constructivists’, they were simply more concerned with the consequences and wished to establish firm grounds for certain knowledge.

iii) The methodological positions were plural; with some positivists being much more flexible in terms of what could count as empirical data or ‘evidence’. They shared one

\(^{22}\) Sometimes positivism is caricatured to such an extent that one wonders how so many acknowledged geniuses could be so stupid. We shouldn’t be so mislead.

\(^{23}\) These were all European-educated, post- Kantian, post-Hegelian philosophers, who were well aware of the problems with naïve realism. This is not to suggest that they didn’t reconfigure a more sophisticated dualistic approach, but one would have to be careful to specify just what ‘dualism’ they were ‘committing’ and why it is such a crime in the first place.
common objective: to establish a parsimonious set of criteria that set apart ‘science’ from ‘non-science’, proper empirical inquiry from metaphysical nonsense. This meant that the type of phenomena that was to be avoided was that which was considered purely ‘subjective’, meaning not that it was derived from human understanding, but that it couldn’t be ‘objectified’ through some publicly shared inter-subjective standard. Thus, depending on how optimistic one was about the methods available for standardized assessment and measurement of particular types of phenomena, certain determinations were made about the proper methodological approach. For example, strict behaviourists believed that mental phenomena should be ruled out of court, because there didn’t exist methods for making ‘public’ what were essentially private phenomena. The main worry was that, without proper objectification, ‘ideas’, ‘beliefs’, ‘values’ etc., would undermine the progress of the scientific enterprise. This was a worry articulated explicitly by Emile Durkheim at the very beginnings of the discipline of sociology.

iv) To explore the ethical position of positivism, one has to immediately let go of the idea that positivism advocates a ‘value-free’ attitude to science. In fact, its very emphasis on separating ‘values’ and ‘facts’ in the process of inquiry derives from a deep ethical commitment. It has to be remembered that many of the greatest positivist philosophers were European Jews at the time of the rise of Nazi Germany, with a profound concern that scientific knowledge was being perverted to meet the nationalistic requirements of an emerging, fascistic element in European culture (long pre-dating the actual takeover of power by the Nazis). Their worry was an important one then and still is now. It is the basic philosophical underpinnings of the continuing debates over the ‘validity’ and ‘reliability’

---

24 This need for standardization and objectivity is the true driving force of the development and ascendancy of quantitative methods, not the other way around (cf. Porter, 1995).
criterion in the evidence debate, and in fact, has been a basic concern for Western philosophy *in toto*, since Plato began his infamous generation of footnotes.

**Implications**

These brief notes on positivism’s philosophical positions allows us to understand that the type of worries that underlie the entire apparatus of EBM are constitutional ones that connect the specific development of protocols and methods with a peculiar obsession with the foundations of inter-subjective knowledge claims. One sees this again and again in various debates over methods that revolve around the uncertainty in knowledge provoked by deviating from rigorous design. In essence, the positivist challenge is that it leaves an indelible mark on the evidence debate with its ‘foundationalism’ and its particular solution-set is circumscribed by its epistemological commitments to resolving the ‘scandal of induction’ (Hacking, 2001).

To better understand this pervasive paradigmatic attachment, it is helpful to consider the vexed issue of ‘validity’. In everyday terms, a valid knowledge claim would seem to mean a representation or description of some aspect of the world that is true. However, positivism, because it is essentially an ‘irrealism’, rules such naïve perspectives out. Validity, for positivists, is basically formal and procedural. What guarantees a knowledge claim’s validity is that it was arrived at through agreed upon, standardized procedures, recognized by a community of scientists as correct method. The reason for this approach to validity is firmly anchored in the ontological and epistemological assumptions referred to above. For positivism, what is real is simply that which can be established as valid knowledge through properly established methodological procedures. It should be noticed that this notion is much more deeply rooted than with simply the small, if not totally vanished, minority of social
scientists who would admit to being ‘positivists’. The philosophical cannons of positivism have long ago been vanquished, but the actual way of seeing the world of scientific activity that generated ‘positivism’ is still as strong today as it has ever been, and certainly drives most of what passes for research on health interventions. To challenge this hegemony means to cut at the very roots of its underlying philosophical assumptions. The next two orientations do this to varying degrees and with quite distinct argumentative strategies.

*Constructivism, Phenomenology and Health Promotion’s Reaction*

‘Constructivism’ is not my favourite word to describe anything, never mind a coherent philosophical approach; however, advocates of an influential modeling of social science ‘paradigms’ have used this term and many have followed, so it is useful to at least address this unfortunate misnomer.²⁵ Health Promotion was in its modern origins a set of ideas in search of a theoretical basis outside of the traditional bio-medical and epidemiological disciplines, so it is not surprising that the strong reaction to the challenge of EBM found succour in the social sciences. At the time of this ‘challenge’, one of the influential trends in the social sciences and particularly, sociology was a general anti-positivism that found its most rhetorically persuasive cadences in a mélange of ‘postmodern’ ideas and critiques. One way of attempting to give a more solid philosophical gloss to this trend was to label it ‘constructivism’. For evaluation, this meant abandoning the ideal of objectivity and embracing the ‘multiple constructions of reality’, valuing stakeholders’

²⁵ I am inclined to agree with Ian Hacking, in his wonderful book, *The Social Construction of What?* (1999), that most talk about ‘constructivism’, either for or against, is redundant and much ado about nothing. It should also be said that there are many varieties of theorists who call themselves constructivists, not all of whom adhere to the philosophical positions adumbrated below. Nevertheless, Guba and Lincoln are the exemplars used here for two reasons: 1) Their textbook interpretation of paradigmatic distinctions (Guba & Lincoln, 1994) has become massively influential for social scientists who are not primarily methodologists and so has a profound effect on many fields of inquiry; 2) Specifically, their interpretation is a dominant reference within the field of evaluation, due to the importance of their book, *Fourth Generation Evaluation* (1989), where a constructivist approach to evaluation is outlined and defended.
knowledge as equal to the evaluation/scientific expert. This rhetoric has more than superficial appeal to health promotion, with its commitment to the principles of empowerment and participation. In fact, some of the language and argument found in the work of constructivism’s defenders mirrors the kind of language in the core documents of health promotion itself.\(^26\) However, as is argued below, philosophically, constructivism shares more of positivism’s basic assumptions than it would like to acknowledge. The philosophical underpinnings of many of the postmodernist/constructivist ideas can be found in a particular interpretation of the phenomenological and hermeneutic traditions. However, there are other approaches to social inquiry\(^27\) that also appeal to these philosophical traditions, but have quite different interpretations of them and therefore come to quite different ontological, epistemological, methodological and ethical conclusions

i) *Ontologically*, most strong constructivists share an irrealist or even anti-realist ontology with the empiricist-positivists.\(^28\) Like positivists, they have no time for the idea that there is a ‘mind-independent’ world with which we can confront our ‘constructions’ of reality. They see this, like consistent positivists did, as a question not possible to answer sensibly. It should be pointed out that this is entirely distinct from the position within branches of both phenomenological and hermeneutic philosophy (branches which include key figures, such as Merleau-Ponty and Gadamer) (Gadamer, 1989; Merleau-Ponty, 1962; Taylor, 2002, 2005). As will become apparent, constructivists underestimate the ontological

---

\(^26\) Not surprisingly, as many of the important innovators in the field, were sympathetic to post- and even anti-positivist social science (e.g. Ilona Kickbusch, Ron Labonte, David McQueen).

\(^27\) I would include both Participative Inquiry (Heron & Reason, 1997) and Ethnomethodology as examples of social inquiry that grow from the phenomenological tradition, yet reach very different conclusions than their radical constructionist counterparts.

\(^28\) As Schwandt (1994) points out in his excellent article on constructivism and interpretivism, the philosopher Nelson Goodman explicitly labelled his constructivist approach ‘irrealist’. While few ‘constructivists’ refer to Goodman’s work, his presentation of an explicitly irrealist ontological perspective corresponds closely to what strong or radical constructivism implicitly defends.
chasm between themselves and other approaches they would like to see as ‘convergent’
(particularly many influenced by phenomenology and participatory research) and
overestimate that chasm between positivism and themselves. For many phenomenologists
and hermeneutic philosophers, the notion that the world can be reduced to the result of our
collective and individual symbolic constructions is a misunderstanding of, forgetting of, or
simply ignorance of, the core concept of the *lebenswelt*. The ‘living-world’ that confronts us
in our interaction with it is understood as antecedent to the subject-object distinction; thus,
locating ‘versions’ of reality entirely on the side of the ‘subjective’ end of the spectrum
simply makes no sense. As Charles Taylor says, in a recent explication of Merleau-Ponty’s
philosophy, “we are only able to form conceptual beliefs guided by our surroundings because
we live in a pre-conceptual engagement with these surroundings, which involves
understanding” (2005, p. 38). As Taylor would argue, constructivism is still caught within
the ‘representationalist construal’ of an ‘inside’ and ‘outside’, where, in their case the
‘outside’ is totally inaccessible. Here foundationalism and relativism are the flip sides of the
same coin, bitter enemies that dialectically reinforce their shared representationalist picture.

ii) *Epistemologically*, the consequences of the constructivist picture of the world as
simply many different versions of reality as constructed by human beings, leads to an anti-
epistemological epistemology, where we cannot know the ‘truth’ of any particular version of
the world in relation to how the world actually is. It also tends to undermine any strong
criteria to judge between the relative adequacy of different ‘constructions’. Although there
are some loose criteria of judgment offered in place of the traditional ones (e.g. Guba and
Lincoln’s), most constructivists have bowed out of the ‘criteriology’ game (Smith & Deemer,
2000). Alternative approaches, such as Participative Inquiry and Ethnomethodology have a
different understanding of epistemological issues. Participative Inquiry, does not abandon the use of terms such as ‘truth’ and ‘validity’, but re-conceptualizes them within an interactive, participatory frame, that understands knowledge as multi-dimensional, with ‘practical knowledge’ as the primary base upon which human beings engage with the world and learn about it. Ethnomethodology, ‘re-specifies’ epistemological questions, by investigating how, in real-world settings, actors decide upon and adjudicate questions of ‘truth’, ‘adequacy’, ‘validity’, ‘correctness’, etc. This latter approach is more interested in the actual purchase ‘epistemic’ issues have in everyday-life (including in laboratory science) and how members of those practices go about solving them for all practical purposes.

iii) Methodologically, constructivists tend to be pluralists, guided by ‘hermeneutic/dialectic methodology’ with a strong penchant for choosing questions that require a variety of ‘qualitative’ methods. There are, in principle, no injunctions against quantitative methods, but in practice, they are seldom used. The approach is aimed mainly at understanding the ‘meaning’ or ‘sense-making’ constructions involved in human interaction. The implications for inquiry are only loosely defined, as the link between theoretical approach, method and data is underdetermined. Interestingly, for an approach that relies on such a radical philosophical stance, constructivism is quite ecumenical when it comes to method choice within the ‘traditional’ qualitative cornucopia; there does not seem to be any particular guidance against the use of specific types of method. This latter point contrasts starkly with Participative Inquiry and Ethnomethodology. Participative Inquiry insists that all methods used to engage in participatory research must themselves be participatory; this rules out many typical quantitative methods, especially surveys, and even some qualitative ones.

29 It is hard to decipher just what this means in terms of specific methodological prescriptions if any.
Ethnomethodology is even more extreme in its approach, with little else countenanced beyond thick, ethnographic description of a very specific sort.

iv) Ethically, constructivism is committed to the “inclusion of participant values in the inquiry” (Guba & Lincoln, 1994, p.115). The ethical stance is that it is important to engage with participants in the inquiry process in order to properly understand their situation and to represent their values and ideas. Participative Inquiry has an explicit aim or goal in all of its work: an axiology of enhancing human flourishing. This approach sees symmetry between the ontological, epistemological, methodological and ethical (axiological).

Ethnomethodology itself has very little reflection upon ‘ethics’ or ‘axiology’ as resources for inquiry, but would rather turn them into topics themselves: what is it that makes social researchers act ‘ethically’? Ethnomethodology tends to systematically undermine the idea that social theorists have a special monopoly on reflective resources and tend to see the notion that a special ‘social science’ perspective is needed here as another form of treating members of social life as ‘cultural dupes’. In other words, you cannot avoid ethical questions in social research because the people you are researching with are just as aware of these problems and will remind you of them if you happen to forget. The question then becomes how do you deal with people ethically, not as a ‘scientist’ but as a member of social life engaging in a particular activity subject to the moral sanction of the community at large.30

Implications

Obviously, this so-called ‘orientation’ is actually internally divided, and one could easily argue that each perspective deserves its own treatment, as would undoubtedly be the case if this was simply a non-specific methodological analysis. However, despite their

30 Many epidemiologists are now finding this out in relation to the ethical demands of the Aboriginal community for research that actually changes something.
differences, none of these approaches directly confronts the EBM solution to synthesizing evidence: how do we arrive at cumulative results from a variety of different studies? It may be fairly said that persons within these approaches have principled reasons for avoiding this issue; however, for my purpose here, such a position is not relevant.

For constructivists, such as Guba and Lincoln, their epistemological relativism and ontological anti-realism makes the idea of review syntheses anathema. A constructivist’s attachment to even a weak version of incommensurability makes it impossible to integrate findings developed from opposing paradigmatic and methodological approaches (Mays, Pope & Popay, 2005). Neither Participatory Inquiry, nor Ethnomethodology have taken up the whole issue of synthesizing evidence and do not promise to do so, as this type of inquiry does not fit with either of their general approaches. At this point it is useful to say something a bit more about Ethnomethodology in this context. For while it is true that this approach would not take it upon itself to produce a first-order evidence synthesis, the type of epistemic topics that such an endeavour raises is classic grist for the mill of ethnomethodological inquiries.

The Ethnomethodological approach also grows out of the interpretive and phenomenological traditions of thought, and sustains a radical critique of positivism; however, its analysis of positivism’s ills and its particular use of the phenomenological tradition departs markedly from many of its ‘qualitative’ brethren. Ethnomethodology developed its program of research on the basis of Alfred Schutz’s phenomenological critique of Talcott Parsons, with Garfinkel taking some of Schutz’s starting points and elaborating a line of empirical inquiry that wanted to investigate how members of social life actively co-constructed the sense and coherence of their everyday activities. The whole attitude of Ethnomethodology has always been to set aside the type of high-epistemological and
ontological concerns that animate the debate between foundationalists and anti-
foundationalists. In fact, the program of research is partly about re-specified the traditional
philosophical concern with ‘truth’, ‘correctness’, ‘validity’, etc., as epistemic topics for
members of everyday life. In other words, the question of whether language can adequately
refer to a ‘reality’ outside of it is moot.\textsuperscript{31}

The problem of developing an alternative methodology for research synthesis can
thus become a ‘topic’ for ethnomethodological analysis in that the attempt to establish
methods and make sense of issues like ‘validity’, ‘reliability’, ‘truth’ in relation to knowledge
synthesis become practical, everyday problems in the context of trying to ‘do’ these
syntheses. The ethnomethodological approach offers a rigorous form of analysis to explicate
the mundane enormity of trying to actually produce an alternative methodological approach
that ‘works on the ground’, so to speak. It is obvious that the initial work on the exigencies of
the much more well-established ‘EBM’ process of doing review syntheses and creating
guidelines is itself a fascinating topic (Moreira, 2005). This makes an incipient attempt to
develop an alternative approach an equally fruitful object of analysis and one that could
potentially inform us of both the progress of such an initiative, but also provide insight into
the generic properties necessary for any actual implementation of such theoretical programs.

\textit{‘Post-positivism’, Critical Realism and Realist Synthesis}

Post-positivism will not be the primary focus of this section, mainly because it is an
umbrella term for almost everything that happened in the philosophy of science after the

crash of positivism. The main focus will be on disentangling the misinterpretation of critical

\textsuperscript{31} Moot because, from the perspective of members of everyday life, this simply doesn’t arise as a concern. The
‘adequacy’ of language as a representational tool is taken for granted, and in fact, in everyday life, people who
raise this as an issue are rightly seen to be either ‘joking’, ‘absurd’, or if pushed too far, even ‘crazy’.
realism that has been spread by Lincoln and Guba's influential lumping of the latter under the 'post-positivist' paradigm. There were many ways and means considered to 'rescue' positivism (and therefore the empiricist tradition) from its demise, including most influentially, Popper's falsificationism, along with Quine's many and varied lines of inquiry. However, to characterize critical realism as 'post-positivist' is to radically misunderstand the history of modern philosophy. First, realism as a philosophical position, pre-dates positivism and is primordially opposed to both positivism's central tenets, and originally to positivism's empiricist heritage. It is unequivocally an 'anti-positivism'. Thus to understand how critical realism offers an alternative to positivist approaches to evaluating effectiveness, we first have to clear up some of the philosophical underbrush.

i) *Ontologically*, critical realism exposit a stratified metaphysics, wherein the world is made up of independent embedded structures and mechanisms, which we as humans interact with and are affected and effected by. Crucially, it believes in the rationality of postulating unobservable entities and has a severe critique of empiricism for ruling such entities out of consideration. It has a three-layered ontology, with the basic layer being the 'real' (this includes all causal powers, potential and actual, liabilities of which are located in mechanisms and structures\(^{32}\)); the second layer is the 'actual' (this includes all actual events as they occur through the contingent interactions of mechanisms and structures); finally, there is the layer of the 'empirical' (these are all those events which are apprehended by human knowledge). Realists accuse empiricists of committing what they call the epistemological fallacy, where ontology is collapsed or 'flattened' into the epistemological or 'empirical' layer. Through a series of transcendental arguments realists postulate what the

\(^{32}\) Some realists deny that 'structures' have causal powers, arguing that only individual mechanisms are causally efficacious.
world must be like in general in order for us to have knowledge of it. Critical realists have also been at the forefront of examining the concept of emergence, where new forms, structures and entities emerge from the non-reductionistic interaction of 'lower' structures and mechanisms. Necessity and contingency are key ontological concepts for realists: necessity is located in the essential aspects or properties of specific mechanisms and structures that give them their causal power; whereas, actual events in the world are the result of the contingent interaction of specific causal chains.

ii) *Epistemologically*, critical realists are 'relativists', in the specific sense that all knowledge claims are, in principle, fallible, due to the fact that knowledge is a human construction, which only ever approximately reaches towards the truth. In addition, critical realists take on the notion that knowledge and its truth-value is 'relative' to the practical nature of the question. It is certainly not true that critical realists are 'probabilists' as Lincoln and Guba claim (2000). 'Probabilism' is a particular solution to an entirely empiricist befuddle: the problem of induction. Realists sidestep this problem because they do not believe that knowledge of something relies on the constant conjunction of empirically observable events.

iii) *Methodologically*, critical realists are also pluralists, believing that the particular methodological approach used should be guided by the specific type of research question that is being asked. In the social sciences, critical realists often advocate a variety of interpretive strategies, in combination with other sometimes-quantitative methods. Seldom do critical

---

33 This is a sort of reversal, or turning on its head, of Kant's transcendental arguments about what human cognition must be like given what we perceive.

34 For example, if you want to know what the chemical structure of water is, $\text{H}_2\text{O}$ is an adequate and accurate answer; however, if you want to know whether a particular stretch of ocean is beautiful or not, a chemical equation is not only inadequate, but absurd.
realists ever advocate for experimental methodologies to guide inquiry in the social world. The most important methodological injunction from a critical realist perspective, is that inquiry should be *explanatory* as well as interpretive, and should aim to explore through both theoretical and empirical investigation, the underlying causal mechanisms that are key to understanding and explaining the reason events happen in the way they do.

iv) *Ethically*, critical realism sees itself as part of the critical theoretical tradition that is axiologically engaged in furthering human emancipation and flourishing. It sees its task as critically examining our knowledge and beliefs, particularly those that uphold the current status quo of human suffering through gross inequalities by denying the capability of transformative, collective human action to make positive change.

The crucial distinction to make in considering a critical realist approach to the issue of assessing effectiveness of health promotion is in its conceptualization of causality. Causal and explanatory language has, within the debate over how to ‘measure’ effectiveness, always been the exclusive preserve of the positivist-empiricist approach; even the most searing critiques of the latter approach have often proceeded by attempting to ‘out flank’ the issue of causality. That is, they have argued, for a variety of reasons, that complex human interactions of the sort involved in health promotion activities should not (exclusively) use the model of causation advocated by the proponents of the randomized controlled trial (RCT). Instead, what is proposed is no model of causation at all, but rather a commitment to ‘understanding’ complex processes, hoping to derive lessons from the accumulation of evaluation results (Nutbeam, 2000). Characteristically, what realists want to say is that, on the one hand such a reaction cedes too much to the positivist-empiricist position on causality, while on the other
hand it hamstrings ‘qualitative’ research by ignoring its significant potential contribution to the development of causal-theoretical conceptualizations.

**Critical Realism and Causality**

Critical realism, as developed by Bhaskar and others, derives from an argument that the actual progress of science does not follow the course of empiricist associational logic, but is much closer to the more traditional realist\(^{35}\) trajectory, where causal relations are postulated and tested based on close study of specific phenomena, and that the regular succession of events, is in fact an artificial creation of laboratory science, for the purpose of isolating real causal mechanisms from the complexity of their contexts. Importantly, the level of confidence in causal claims does not rely on multiple RCTs, but on the theoretical context within which causal claims are made. Bhaskar’s original analysis, in *A Realist Theory of Science*, was an attempt to account, philosophically, for the way in which scientific experimentation was actually carried out in the natural sciences (1975). Thus, the critical realist critique was aimed at the entire interpretation of science as it had been handed down by the empiricist tradition in the philosophy of science. Bhaskar argued that the empiricist understanding of causality could not logically account for the way experimental science was practiced\(^{36}\). First, there would be no point in experimental activity if the world was structured causally into regular conjunctions of events (i.e. B always following A); the point of experimental activity is precisely to isolate the causal ingredients one is interested in order to create a regular pattern of events and confirm one’s causal intuitions about natural laws. Conversely, if causal relations only pertained to regular successions of events, then the

\(^{35}\) Bhaskar himself was profoundly influenced by the work of Rom Harre (Harre, 1970; 1974; Harre & Secord, 1972). Harre is perhaps the most influential modern realist particularly in the British context.

\(^{36}\) Bhaskar consciously chooses the strongest basis of the empiricist position to critique.
results of experimental activity (the identification of natural laws) would be pointless, because the isolated causal ingredients, once ‘put back into nature’, would no longer have causal effect. The crucial point of this logical argument is that, it is an inevitable pre-supposition of scientific activity that the mechanisms isolated in experimental conditions, continue to act causally in nature, even though, empirically (that is, through natural observation), they may not have the predicted effect. The classic example is the law of gravity. The law of gravity has specific causal powers, whether or not for any specific case it produces observable effects. So, an airplane does not fall to the ground, not because the law of gravity ceases to work, but because through counter-acting causal forces, it can continue to fly, despite gravitational force.

The realist case relies on the argument that, in fact, science does not work according to the empiricist associational logic. As was pointed out earlier, the ‘biologic credibility’ of an epidemiological hypothesis often rests on scientific discoveries that never came about in the way one would expect, given the empiricist account of science. For example, William Harvey’s discovery of blood circulation and his critique of Galen relied upon creative experimentation and arguments that showed why the previous non-circulatory theories could not explain a series of important phenomena (Porter, 1997); obviously, it being the early 17th Century, he did not have access to modern statistics, yet his discovery is still ‘biologically credible’. Scientific laws (the pinnacle of scientific achievement) are not discovered through the employment of pure associational logic, but through the identification of real causal mechanisms (such as the discovery of DNA). Once this argument is accepted, it is clear that the process of discovering causal mechanisms has much more to do with the careful conceptualization of phenomena, than with the analysis of a corpus of statistical data. It is an
important argument that distinguishes it from much of the interpretive tradition of social inquiry, in that it does not accept the empiricist interpretation of natural science or of causality.

As Andrew Sayer says: “To ask for the cause of something is to ask what ‘makes it happen’, what ‘produces’, ‘generates’, ‘creates’, or ‘determines’ it, or, more weakly, what ‘enables’ or ‘leads to’ it” (1992, p.104). As one can see, this is an entirely different approach to the analysis of causality. It is not about the relationship between discrete events (how many times B follows A), but about how certain objects or relations have ‘causal powers’ or ‘liabilities’ or tendencies to act in a certain way. Identifying a causal mechanism is to identify how some object or relation ‘works’; it is to analyse or abstract the essential attributes of a thing in order to understand its “way of acting” (Sayer, 1992, p.105). This identification does not rely on the object or relation’s effects in any particular pattern of events. For example, identifying the ‘powers’ of the prime minister does not depend on whether or not these ‘powers’ are actually activated. We know that the prime minister, if he or she wants, can fire the entire cabinet; this is an inherent ‘causal power’ of the prime ministerial office. Whether such a power is activated or not depends on the contingent conditions or ‘context’ within which those powers are situated. Thus, a causal claim “is not about a regularity between separate things or events but about what an object is like and what it can do and only derivatively what it will do in any particular situation” (p.105).

Implications

The implications of a realist approach to systematic review have been most articulately laid out by Ray Pawson (2002b). The key aspect of Pawson’s critique of the two prevailing approaches to systematic review (meta-analysis and narrative review) is that both
are caught up in trying to identify programs or initiatives that work, rather than focusing on what key mechanisms are responsible for the way in which any particular initiative unfolds. For Pawson, this means developing a method that takes into account the key mechanisms (M) thought to be causally related to successful outcomes, the particular contexts (C) in terms of people, resources, and the local environments with which these key mechanisms interact, and the particular outcome pattern (O) that is empirically observable. The twin objectives of this realist approach to synthesis are to identify the mechanisms that are the drivers of successful change and to identify and categorize the core circumstances that ‘blunt’ the effectiveness of particular mechanisms. For example, involving key stakeholders has been identified in the literature as a key, indeed necessary mechanism for successful community interventions; yet, what are the factors that make such involvement difficult? The contexts within which different implementations of policy mechanisms are practiced can be abstractly defined and compared with the outcomes of those implementations. The important point here is to recognize the need and the possibility of a ‘middle-level’ of analysis, where context is taken into account, but a certain level of abstraction is accepted as the foundation for broader comparisons. It is true that no context of implementation is identical to any other; yet, the same could be said for any situation in the world at all. For informative description to be possible, never mind adequate explanation, abstraction is just unavoidable. As Sayer points out: “To be practically adequate, knowledge must grasp the differentiations of the world; we need a way of individuating objects, and of characterizing their attributes and relationships. To be adequate for a specific purpose it must abstract from particular conditions, excluding those which have no significant effect in order to focus on those which
do. Even where we are interested in wholes we must select and abstract their constituents” (1992, p. 86).

In the literature on community interventions there is some strong evidence to support the candidate mechanisms of successful action and outcomes. Nevertheless, there is not nearly enough evaluation data that describes processes in detail in the published literature. When it comes to defining and categorizing contexts that support these mechanisms and contexts that blunt their effectivity, there is of course, even less data to rely upon. The realist approach in this particular area, must reconstruct from the existing literature, contextual analyses, and encourage actively, future evaluation work that focuses on the compilation of relevant data. The argument here is threefold: first, existing approaches have been inadequate, and so an alternative is definitely needed; second, there is a logical framework offered, that, if it had good evaluation data to rely upon, could offer strong indicators of the overall effectiveness of community initiatives and distinguish between more and less effective approaches; third, and perhaps most importantly, an approach that actively encourages the collection of data that people care about and believe is at the core of their vision, is much more likely to produce quality evaluation reports.

Summary

What has become abundantly clear after a review of the nascent attempts to build a solid evidence base for health promotion through the use of various forms of knowledge synthesis, mostly based on the established protocols emanating from EBM, is that alternative approaches to producing evidence and then synthesizing it are desperately needed. However, simply because they are desperately needed does not mean they will be successful. We can see highlighted in the analysis of some of the deep philosophical issues surrounding this
problematic, that whatever path is taken in this quest, it will not be easy. Nevertheless, all alternative paradigms and all alternative theoretical resources need to start somewhere, and all dominant paradigms started off in equally marginalized positions, with little prospects for success. EBM itself, now seen by many as a hegemonic bastion of 'scientific truth' was not long ago the wild idea of a maverick epidemiologist (Cochrane, 1972). Furthermore, EBM rests on the more or less 'solid' foundation generated by over two centuries of progress in statistical and epidemiological innovation. Following Imre Lakatos (1977), what is really needed is an appreciation of whether any of the barely conceived, embryological alternatives have the potential to generate a serious 'program of research'. Of course, none of the existing candidates, including 'realist synthesis', are far enough along for us to be confident in such a prospective judgment, yet many interesting insights are possible as we watch their struggle to be born.

Research Question

By 'research program', I mean to invoke Lakatos' concept of 'progressive' or 'degenerative' research programs, without borrowing wholesale his philosophical framework. In particular, the intention is to put some sociological flesh on the philosophical bones that make up Lakatos' 'programs'. Rather than just sets of interrelated hypotheses, research programs will be defined here as: a relatively coherent body of philosophical commitments, theoretical concepts and hypotheses, methodological strategies and empirical methods, that are embedded in practical, social, political, economic and cultural contexts, and are aimed at answering a relatively coherent set of important research questions.

For research programs to progress, they not only have to 'discover novel facts' (although, this is indeed important), they have to, among many other things, attract
experienced researchers, new investigators and graduate students; they have to be able to successfully train these new researchers and graduate students in the practical research skills and tools necessary to carry out the research; as well, eventually, progressive programs begin to see their ‘findings’ as part of established curricula being disseminated and taught across the disciplinary community. At some points, certain ‘research programs’ become so well established that they approach what Kuhn referred to as ‘paradigms’, or ‘the way we do things around here’, as one famous philosopher of science put it provocatively (Rorty, 1989). So what hope do alternatives to the RCT and the whole EBM approach have for assessing the effectiveness of health promotion? Are some alternatives more promising than others? What type of criteria would we want to use to make such judgments? Are these approaches still ‘scientific’? Or, are these quite impossible questions to answer sensibly, or even, simply wrongheaded ones?

To approach this potential morass in an appropriately deflationary way, I have decided to take one of the potential alternative approaches to reviewing evidence of health promotion effectiveness as a case study: the realist synthesis approach. I have done this in order to explore some of these questions more deeply, but also to bring a more concrete focus to the mundane, everyday aspects of ‘doing a research program’.

The main question that this study aims to answer is: what are the lebenswelt foundations of a realist synthesis approach? That is, in the actual lived course of doing a realist synthesis, what are some of the key features that become apparent as necessary foundations of doing this work. We already know much about realism’s ontology and epistemology (this is reviewed above); what about the in-the-world epistemic and ontic practices of implementing the realist synthesis recipe. The argument here is that to truly
address the 'viability' of the realist alternative (or any alternative, including the EBM orthodoxy), this question is just as important as any other, and in fact, it is a question that is likely to be systematically neglected otherwise, for reasons explicated in the next chapter.
CHAPTER 3

METHODOLOGY AND METHODS:

An Ethnomethodological Approach

to the Sociology of Social Scientific Knowledge

In terms of philosophical and historical context, I have drawn on work that has been taking place in the contemporary history of science. Based primarily on Theodore Porter (1986, 1995) and Ian Hacking’s (1990, 2001, 2006) work, along with a whole network and field of research (Buchwald, 1995; Wise, 1995) in this area. There is now an emergent, philosophically informed historical analysis, that gives a clearer picture of what lay behind the great flowering of the positivistic, statistical, quantitative social sciences over the past two centuries. This analysis is crucial for understanding the context that health promotion finds itself in concerning the ‘evidence debate’, and therefore, it has been drawn upon in various ways, and many of the assumptions of this work is critically adapted, to paint the background of the situation within which health promotion now finds itself.

However, as a direct, first order empirical inquiry, this study can be thought of as one contribution to what Maynard and Shaeffer (2000) have referred to as a ‘Sociology of Social Scientific Knowledge’ (SSSK). This picks up on the large body of work done in the areas of Social Studies of Science and the Sociology of Scientific Knowledge, which have been mainly focused on sociological investigations of the ‘natural’ sciences. Using many of the insights gained from this work, a SSSK study aims to elucidate the practical, situated nature of conducting ‘social science’ endeavours, particularly those ubiquitous attempts to objectivize the ‘subjective realm’ (Maynard & Schaeffer, 2000).
The specific approach taken here is most heavily influenced by the ethnomethodological studies undertaken by Harold Garfinkel and his followers (Coulter, 1990; Garfinkel, 1967; Heritage, 1984). In particular, it has been the later developments of the ‘studies-of-work’ programme, and specifically, the work carried out as studies of the ‘laboratory sciences’ (Lynch, 1985). Thus, the methodological implications are that to understand what it takes to develop a research project, never mind a research program, one has to take into account more than simply the theoretical ‘gloss’, or even the more detailed ‘guidelines’ produced by advocates of a particular approach; one must attend to the situated details of actually implementing mundane data collection procedures, building and using various representational tools, maps, tables, etc., and applying a multitude of often unacknowledged heuristic analytical devices to do the work of a research project of the type involved in managing the completion of a realist synthesis.

Before outlining in more detail the appropriate methods for data collection that will be used in order to carry out an ethnomethodological analysis, more needs to be said about three of the fundamental influences on ethnomethodology’s origins and its subsequent progress. First, a brief outline will be given of the nature of ethnomethodology’s relationship to the social sciences in general, particularly in relation to Garfinkel’s working out of what he saw as problematic with Talcott Parson’s approach and how he understood these problems as foundational and shared by most social science. Second, an analysis of the relationship between phenomenology and ethnomethodology. These two elements of influence are intimately tied together, as the former critique builds upon the work of the great social phenomenologist, Alfred Schutz, and the former relationship is consistently interpreted in the light of fundamental problems for what Garfinkel likes to now call ‘the worldwide social
science movement’. Nevertheless, they are treated separately here in order to tease out the nature of ethnomethodology’s progress and to better understand the nature of its current status and interests. Finally, a few words about the influence of the later philosophy of Ludwig Wittgenstein will be considered.

An important caveat should be entered into the record at this point. One of Garfinkel’s more (or less) endearing qualities and one of his most frustrating for those wishing to ‘pin him down’ in terms of intellectual heritage is that he has obdurately refused to give special respect to antecedent sources of inspiration. Although he is a dedicated and somewhat excessive acknowledger of historical and contemporary contributors, he has consistently enacted a playful, yet rigorous disrespect for the high church, ritualistic observances of the ‘proper’ understanding of illustrious forebears (Garfinkel, 2007). This needs to be kept in mind, particularly in relation to his conceptualization of phenomenology and the ‘policy’ formulations derived from this rich heritage. Garfinkel, especially in his most recent writings, seems particularly proud of his intentional misreadings of the grand Tradition (2007).

‘Overturning’ Social Science ‘Indifferently’:

*Ethnomethodology as a Radical Critique of ‘Constructive Analysis’*

Much of what follows is indebted to John Heritage’s masterful explication of the origins of Garfinkel’s work in relation to the theoretical approach of Talcott Parsons (1984). It was out of Garfinkel’s engagement with the work of Parsons (his PhD supervisor at Harvard in the late-1940s), that he developed a sustained and rigorous understanding of just what was at stake in the type of sociological theorizing represented by what Garfinkel would later term ‘constructive analysis’. Garfinkel’s approach was to utilize the work of Schutz and
Gurwitsch to understand the central issues and problems surrounding the sociological theory of action; an area in which Parsons was the leading figure of his generation and an area about which Parsons had fashioned a remarkable theoretical synthesis in his The Structure of Social Action (1937).

Talcott Parsons was profoundly influenced by the Weberian critique of utilitarianism as an approach that subordinated values to the instrumental-rational sphere and understood social action and therefore social structure as being determined by the collective working out of individual self-interest, where restraining social institutions were developed out of an ‘enlightening’ of cruder and less strategic desires these individuals may have. Parsons shared with the Weber and the neo-Kantians the view that social values were more fundamental aspects of social life and were not mere epiphenomena of the clash of individual self-interests. Parson’s solution was to develop what he called the ‘voluntaristic theory of action’. This approach retained Weber’s Verstehende concept of subjectively meaningful normative aspects of conduct, while moving toward a synthesis with a systems approach which integrated an account of the restraining conditions of action which confronted social actors in their efforts to realize subjectively meaningful goals. However, it is the particular approach Parsons took to conceptualizing the place of social norms as part of the constraining elements of the social system, that drove Garfinkel, following Schutz, to question the entire apparatus of Parsonian theory. One of the key things to keep in mind when trying to understand ethnmethodology’s persistent and seemingly obstinate reluctance to associate with mainstream sociology, and its practiced indifference to what are seen as the ‘canonical’ concerns of the discipline, is that ethnmethodologists, following Garfinkel, had and still have, a deep respect for Parsons for his extraordinary accomplishment. This sounds like a strange idea:
how can they make Parsons the fulcrum and prime target of their foundational critique, while at the same time laud his project as a wonderful piece of sociological theorizing? Yet, to understand this paradox is to move some way toward understanding ethnomethodology itself.

For Garfinkel, Parsons was the ultimate sociological theorist. There is reason to take this evaluation seriously. It is hard to find any contemporary theorist, other than perhaps the extraordinary Niklas Luhmann[^37], that even approaches the level of consistent, rigorous, systematic theory building, exemplified by Parsons. This has lead to two different conclusions. First, there is not a lot of patience within ethnomethodology for claims that Parsons has been surpassed by newer and better theorizing, particularly when the quality of the latest efforts pale in comparison to Parson’s great enterprise. More important though is that Garfinkel was convinced that Parsons had dug down to the very core of the fundamental question for social science: what place does the self-understanding of social actors have in relation to the “underpinnings of their own actions” (Heritage, 1984: 30). For Garfinkel, the classic and pervasive answer that social scientists have to this question is that the explanation of social action cannot rely on the resources members of social life have themselves and must rely on external resources that only the social scientist has access to. Garfinkel fundamentally rejects this answer. As Heritage puts it:

“…Garfinkel rejected absolutely the view that the ordinary judgements of mundane social actors can in any way, or under any circumstances, be treated as irrelevant or epiphenomenal in the analysis of social action or social organization. It is the Parsonian disregard for the entire common-sense world in which ordinary actors choose courses of action on the basis of detailed practical considerations and judgements which are intelligible and accountable to others, which ultimately constitutes the central focus and point of departure for Garfinkel’s treatment of the theory of action” (1984: 34).

[^37]: Like Parsons, Luhmann has been the subject of important critiques (Habermas, 1996). However, one thing he could not be accused of is lackadaisical theorizing. His approach to systems theory is dauntingly self-consistent, yet, as one commentator puts it “forbiddingly dense” (Lechner, 2000).
This rejection has fateful consequences for any and all social science projects that turn on producing extrinsic accounts of social action. This perceived rejectionism has led to many confusions about Garfinkel’s project. We should perhaps go through a few of them in order to avoid dealing (yet again) with pervasive misperceptions. A typical response to Garfinkel’s ‘rejection’ is to argue that such an approach acts as if social actors ‘can never be wrong’, as if whatever understanding members have of their activities is all that can be said about them: thus, the story goes, according to Garfinkel and people like him (Peter Winch is a favourite target), we end with the absurd conclusion that the world is just as people conceive of it and that cognitive error is impossible. When one reads this type of facile argument, one wonders whether much thought has really gone into how such a smart person (along with his followers of course) could be so stupid. The fact is, that neither Garfinkel, nor any of his followers, have ever held such an absurd view. In fact, for anyone that bothers to pay attention for more than a few minutes to the output of ethnomethodological work (and in fact, to any and every walk of social life), one of the most notable and ubiquitous aspects of human interaction is its constant topicalizing of ‘correctness’, ‘error’, ‘truth’, what is only ‘apparent’, ‘what may be the case’, ‘what we are certain of’, ‘what is doubtful’, etc, etc. People are constantly aware of the possibility of errors in judgments, lapses in reasoning, jumping to conclusions, overlooking the obvious, forgetting something essential, etc. Furthermore, ‘people’, as members in everyday activities, constantly hold each other to account just for these types of lapses. What Garfinkel denies is that members of social life can be somehow absolutely and universally ignorant of some special insight or insights about the nature of their own activities and that only special social scientific methods can remedy this situation. In other words, he rejects the notion that the explanation for the outcome of
social action can be supplied by an exclusively extrinsic, constructive analytical framework, from which we ‘read off’, retrospectively, people’s behaviour.

Another easily made misperception is that ethnomethodology is a kind of ‘ironic’ dismissal of social scientific activity as a whole. This view is held despite consistent warnings to the contrary (Garfinkel, 1967; Heritage, 1984; Sharrock, 1989). The attitude of ‘ethnomethodological indifference’ has driven many commentators to despair as they try vainly to figure out just what metaphysical, epistemological or ontological position ethnomethodology is taking. However, in this sense ethnomethodology is not taking one ‘position’ among others in the grand social theory debate. On the contrary, it disavows this project as simply not its ‘interest’, and plows along investigating “the organizational relevances that inhabitants of the society treat as the actual conditions of their conduct” (Sharrock, 1989: 664). Although ethnomethodology often ‘sounds’ like it is dismissing other forms of social science activity, it is actually more annoyed at the constant requests to repeatedly justify its activity entirely on the terms of ‘constructive’ social science. There is no request on behalf of ethnomethodology for other approaches to social analysis to cease and desist from their activities; these activities are just another worldly practice open to ethnomethodological analysis. Often the core issue here revolves around the perturbing fact for orthodox ‘constructive analysis’ that their programmatic activities are essentially impossible to pursue consistently, without relying on exactly those everyday, mundane forms of reasoning that they set out to try and eliminate from their theoretical programs. This is not, it must be said an accusation hurled by ethnomethodologists; it is a pervasive topic of all serious methodological self-reflection in the social sciences. The most scathing critiques of the incomplete, inadequate and primitive state of theory, data and measurement in the social
science disciplines come not from ethnomethodologists, but from the most brilliant and
senior methodologists in those disciplines. As Sharrock points out:

"...most of the argument that is collected under the heading 'the
ethnomethodological critique' does not comprise an enumeration of faults
which ethnomethodologists have found but which others had previously failed
to appreciate. The critique does not, at an important level, provide news to
conventional sociologists about their theoretical and methodological practice,
but mentions, rather, matters of which they are—at least in practice—only to
well aware" (1989: 666).

This leaves us with the perspectival attitude switch taken by ethnomethodology. Most
social science treats its chronic methodological troubles as grist for further inquiry and
theoretical development, with the hope that at some point in time, it will not have to rely on a
kind of ad hoc, 'for all practical purposes', reasoning in order to repair its insufficient
methodological prescriptions and fulfill its programmatic objectives. For ethnomethodology,
it takes this reliance, that is admittedly pervasive and seemingly unremediable (at least for
now), as a topic for inquiry. It is this perspectival shift that this study takes in order to treat
'indifferently' the question of whether 'realist synthesis' is in fact a valid project, and to
instead focus on the type of practical reasoning that it must rely on in order to make it a
defensible, accountable activity: in other words, it is the perspective necessary to answer, in
an honest way, (i.e. in a way that doesn't sweep its troubles under the carpet) just what is
necessary for realist synthesis to 'work'.

Taking Phenomenology Seriously:

Garfinkel's Reading of the Phenomenological Tradition

The title of this section is somewhat ironic because, as Garfinkel makes clear in
his most recent published work (a rare event indeed), he has little time for reading
Husserl, for example, in any sort of conventional scholarly fashion. He asks:
“What is the relationship between Husserl’s Propositional Program and Ethnomethodology’s Studies of Lived Work and Sciences? How is the relationship between Husserl’s Programme to Specify the Lebenswelt Origins of the Sciences and Ethnomethodology’s Studies of Work and Science to be understood?” (2007: 10).

He then goes on to explain how he will co-author with Andy Crabtree, a ‘Motivated Book Review’ of Husserl’s *The Idea of Phenomenology*, that will “make deliberately indispensable but respectful unfavorable comparisons of Husserl’s program of studies of lived work in the sciences with Ethnomethodology’s program” (ibid.) Because this promised book review is yet to come out, we can only speculate what these ‘respectful unfavorable comparisons’ might be. However, we can give some sense of what type of departure point ethnomethodology makes from the ‘Phenomenological Tradition’.

Phenomenology proper certainly began with Edmund Husserl. Although many philosophers before Husserl used the term (including Kant and Hegel), Husserl was the one to turn it into an entire program and *method* of philosophizing (Merleau-Ponty, 1962; Moran, 2000). Husserl’s famous injunction to ‘return to the things themselves’ was motivated by the felt need to provide an ultimate grounding and foundation for knowledge through the careful description of the essential aspects of phenomena as they presented themselves intuitively to consciousness. Crucial to Husserl’s program was that *description* must come before *explanation*. In fact, the latter was ineradicably parasitic upon the former, and thus any contamination of explanatory schema, including that provided by science, would occlude the essence of phenomena. Initially, Husserl described what he took to be the foundational origins of logic and geometry in the living world of conscious experience. He subsequently made many different attempts to approach what he had set up as his fundamental problem: how to carefully describe *things* as they appear to consciousness. The development of Husserl’s project lead him further and further toward an analysis of ‘transcendental
subjectivity’, with the felt need to ground all pure, a priori knowledge, in a transcendental ego. Without having the space to give a full treatment of the complexity of Husserl’s project, we can still say that, however transformed by later thinkers, his work has had a lasting impact on ethnomethodology.

Husserl initiates an attitude of a ‘ground up’ approach to knowing. This meant that we must start with describing the most basic elements of our conscious experience. For Husserl, his starting point happened to be the ideal essences of the objects of mathematics and logic. Throughout his development Husserl always kept to a kind of foundationalism, where ultimately, the most important knowledge was that grounded in the realm of these a priori objects. Later in his development, Husserl outlined a full-blown transcendental idealism, attached to these ideas about ‘essential seeing’ (Wesenserschauung) (Husserl, 1931). Later, however, particularly in the Crisis in the European Sciences (1970), Husserl paid more attention to the more complex, richly populated, lebenswelt. It is this focus on the investigation of the everyday life-world and the ‘natural attitude’ that was taken up by the two phenomenologists that have had the most impact on ethnomethodology: Alfred Schutz and Maurice Merleau-Ponty.

Alfred Schutz was certainly the original inspiration for Garfinkel’s critique of Parsons and in fact, as a contemporary colleague of Parsons, developed his own direct critical confrontation with the Parsonian project. For Schutz, Parson’s approach could not account for the plurality of ‘rationalities’ that exist in the social world. For, Schutz a phenomenological description of the social world as it presented itself to us a many-layered structure of typifications that were carried out by members of social life themselves. In other words, unlike Weberian ‘ideal types’, which were the product of scientific inquiry, Schutz’s
typifications were the result of the concerted activity of members of social life as they co-
constituted the sense and coherence of their practical activities. For Schutz, his inquiries led
him to develop a series of analyses that provided a phenomenological ‘mapping’ of what he
called the “finite symbolic provinces of meaning” (Schutz, 1967). For Schutz, these
‘provinces of meaning’ were different phenomenological ‘worlds’ such as ‘theatre’, ‘art’,
‘religion’, ‘fantasy’, ‘dreams’ and even scientific theory (Vaitkus, 2000). These ‘worlds’
were differentiated from the ‘natural attitude’ of everyday life, in that they, in various ways,
‘shook’ us out of our taken for granted assumptions and gave us a different set of coordinates
around which to make sense of what was ‘going on’. For Garfinkel, Schutz’s work provided
a key departure point for considering how members of social life actively co-constructed the
sense and coherence of their everyday worlds. However, Garfinkel was concerned to move
away from the more speculative, introspective philosophizing that led Schutz to ‘divide’ up
the world into particular ‘levels’ and ‘provinces’, and to provide a strictly ‘empirical’
analysis of just how, in specific ‘found’ activities, members of social life went about doing
their work of reflexively accounting for what it was they were doing as concerted action.

Merleau-Ponty is important for other reasons, and has mostly been influential at the
periphery of ethnomethodology’s development. His influence was first made most explicit in
the work of David Sudnow (1978), in his fascinating explication of the embodied practice of
improvisation in jazz piano playing. As a practicing jazz musician, Sudnow paid close
attention to the ‘lived body’ and how it comported itself in relation to the concerted spatio-
temporality of collaborative jazz improvisation. Merleau-Ponty’s great contribution was to
radically ‘de-transecndentalize’ and ‘de-idealize’ the phenomenological subject. Although he
adhered to some form of transcendental argument, Merleau-Ponty took up Husserl’s later
injunction to pay attention to the structures of the *lebenswelt* and developed his own rigorous analysis of how the ‘lived body’ (*Lieb*) actively engaged in the world and reflexively constructed the sense and coherence of phenomenological perception. This attention to the details of embodied practice has left a profound legacy for many in ethnomethodology, especially in studies of work, where bodily comportment and orientation is a key feature of interactive sense-making, and where the sub-lingual communicative practice of ‘gesture’ can encompasses a wealth of interactive resources. Recently, other ethnomethodologists have been paying more attention to the influential work of Merleau-Ponty (Crabtree, 2001). Perhaps, most importantly, Merleau-Ponty re-emphasized the key feature of phenomenological work: that it is, and must be, an unceasingly re-invented practice of philosophizing, *not* a set of prescriptive epistemological positions, ontological commitments, or methodological criteria. This *attitude* is certainly the one taken up by Garfinkel and his followers. It is one of the more frustrating issues for outside commentators: there are no set methodological frameworks or guidelines in ethnomethodology. Rather, what has developed is a set of ‘policies’ which are more or less fashioned in order to constantly bring the attention back to the typical pitfalls that bedevil social science practice. For example, there is much talk about what Garfinkel calls the ‘unique adequacy requirement’ (1967), which call for the analyst of a specific social practice to be knowledgeable enough of that practice ‘from the inside’, that is, how members understand it. This has lead some to fear that Garfinkel is prescribing that only fully-fledged members of a social practice or work setting can give adequate descriptions of that social practice or work setting. However, this is to miss the point. The real issue is that most social science actively *discounts* and even denigrates the understanding members have of their own practices in order to substitute a ‘more adequate’
social scientific account of ‘what is really going on’. Garfinkel’s policy here is simply an admonition and mnemonic device for us to pay attention to the perspective that undermines this typical social science approach and elevates the member’s understanding and knowledge back to its rightful place.

Finally, we consider the role of Wittgenstein for ethnomethodology’s progress. In his later philosophy, particularly in his *Philosophical Investigations* (1958), Wittgenstein laid out an uncompromising demolition of what he called the ‘picture theory’ of language. Wittgenstein, as perhaps the most brilliant expositor of a formal analysis of language ever to make the attempt, in his *Tractatus Logico-Philosophicus* (1961), made an about turn and repudiated the entire project as based on a fundamental mistake concerning the nature of language. He offered instead an approach which was tied fundamentally to the actual *use* that members of society made of language, in what he famously called ‘language games’. Of course, Wittgenstein is probably the most misinterpreted philosopher to have ever walked the earth, and there is an almost endless literature that parses the various philosophical disputes that have arisen concerning ‘correct interpretation’ of his works. However, in relation to ethnomethodology, there are two fundamental issues that Wittgenstein has had an important influence on. First, Wittgenstein, more than anyone else, recognized that the so-called ‘problem’ of ‘indexical expressions’³⁸, was actually ubiquitous, general, and rather than a problem, was actually a *resource* for efficient and effective communication. Much of *Philosophical Investigations* is given over to pointing out some of the absurdities that are inevitable if we hold to a representationalist (‘picture theory’) view of language and miss the

³⁸ Indexical expressions are classically indicative expressions such as ‘this’, ‘that’, ‘there’, ‘I’, ‘him’, which, by their very nature, undermine the attempt to give a fully consistent, formal and objective, third person account of language as a system.
way in which the *context* of language use provides for the background knowledge which we rely on to make sense of just what people are saying *as and when* they say it. Many ethnomethodologists have found philosophical arguments in Wittgenstein to support their analogous view that social interaction gets its sense from locally situated contexts, wherein members of social life go about explicating and accounting for the sense of what they are doing.

We can now see only some of the influences that have motivated ethnomethodologists to pay attention to the local details of interaction in order to explicate the sense and coherence of specific social activities. The injunction that we started with from Husserl, to go back to the ‘things themselves’, gets translated in ethnomethodology to urge us to ‘go back to the practices themselves’.

The ‘work’ we are interested in this study involved a series of ‘work sessions’ of a health promotion *effectiveness* project, using a ‘realist synthesis’ approach, with the project coordinator and a small group of research assistants. The sessions were video and audio-recorded and then transcribed. There are four work sessions that were recorded for analysis. The excerpts below are taken from two of the key sessions. Each session was organized around a particular component of the project the team was working on, including: a general discussion about direction and data collection; a discussion about ‘mechanisms’; a discussion about ‘context’; and, a discussion about ‘outcomes’. The analysis chooses particular sequences of interaction in order to elucidate the intrinsic, practical methodological troubles that pervade the attempt to develop and implement the realist synthesis approach in a particular instance. The analysis is tied closely to the data and is an attempt to unpack the unfolding interactive order that organizes the cognitive, affective and even physiological (if
interpreted to mean the ‘lived body’) elements that guide the work of the team as it tries to make sense of the task at hand. Through this analysis, it is shown that, while important theoretical and methodological work is still required in order to develop further such alternative approaches as realist synthesis, there is a pervasive, practical, mundane underpinning, tied to common-sense reasoning and judgment that supports and actually makes accountable the real work of social science and in this specific case, the work of doing a review synthesis.

A rigorous analysis of these sorts of details must rely on the careful recording of the work of doing the research itself.

Methods and Data Collection

Along with secondary sources and historical documents, the primary data collection method that was used in order to support an ethnographical study of an ongoing attempt to implement the realist synthesis approach to assessing the effectiveness of health promotion interventions, along with the wider question of the potential for alternative approaches in general, is ethnomethodological participant observation

Participant Observation

Participant observation is a well-established method in social science (Becker & Geer, 1970). Its early development in the Chicago School of the 1920s and 1930s (see Bulmer, 1984) came out of a concern with elucidating the specific milieu of urban neighbourhoods, particularly new immigrant populations and impoverished communities. It was later that more philosophical and methodological rationales were identified with this method, particularly in the case of Symbolic Interactionism (Bulmer, 1954; 1969). There are many reasons that researchers give for using participant observation, from the use as a sort of
underlabourer for more formal, quantitative research, such as the development of survey
questionnaires, to a much more principled, ‘theory-specific’ method, as Ackroyd and Hughes
put it (1992). As a ‘principled’ choice of method, participant observation is associated with
“a perspective which stresses the interactive, interpretative and negotiated basis of social life,
created and sustained by the meanings actors use to make sense of and interpret the world in
which they live” (1992). The core purpose of the method is to understand a social setting
from the inside out, rather than through the filter of an external interpretive framework. The
claim is that to understand social life, we have to understand the meaningful interaction of
human beings as they go about their daily activities. For this study, an ethnomethodological
approach was taken to participant observation. This means that attention is paid to how
members of the work activity that I will be observing and participating in, account for the
practical production of a local sense of order in trying to make sense of the task at hand: that
is, the accomplishment of a systematic review of health promotion interventions, using the
realist synthesis approach. In particular, as this study sees itself as a SSSK contribution, keen
attention was focused on the variety of materials, rationales and discursive contexts that
make up the work of doing the review. In order to capture this level of detail, particularly the
use of material artefacts, such as charts, graphs, tables, etc., a series of ‘work sessions’ were
video-recorded and transcribed. The refinement of the method of video-recording as a data
collection technique has recently been extended, along with the techniques for transcription
and analysis of video-recordings (Heath & Luff, 1992; Silverman, 2000).
CHAPTER 4
RESULTS AND DISCUSSION

This chapter will address the broad question: what is, practically, at stake in the attempt to implement a specific alternative approach to assessing the effectiveness of community health promotion initiatives aimed at reducing health disparities?

This question requires ethnomethodological analysis of the practical activity involved in doing the actual work of producing a review synthesis of community-based health promotion interventions in North America aimed at reducing health disparities, using a realist synthesis approach. What follows are two sections organized around generic themes that were overwhelmingly apparent as fundamental aspects of the work of the video-recorded sessions. The first section focuses exclusively on the work session on ‘Context’, while the second section focuses exclusively on the session on ‘Outcomes’. These sessions were chosen as the primary foci because they were they ones that most perspicuously pointed to the types of methodological troubles that are endemic to this type of enterprise.

Categorization, Abstraction and the Situated Conceptualization of Context

As has already been alluded to in Chapter 2, there is an ongoing concern with how to integrate ‘context’ into an analysis of health promotion effectiveness. One of the key criticisms of the EBM reliance on RCT evidence is that it excludes attention to contextual issues. As McQueen has recently summarized, “it is part of the very nature of health promotion interventions that they operate in everyday life situations, in a particular context, involving changing aspects of the intervention (2007, p.296)”. There are similarities and even some overlap between the problem of context and the discussion of ‘complexity’ in Chapter 2. There are a variety of levels at which ‘context’ presents difficulties for looking at health
promotion effectiveness. One can consider this problem as one of analysis, measurement, or conceptualization. Often we put most emphasis on the first level, although ironically, this level actually constitutes the finished superstructure of the effectiveness research program. At this level, we become excited about potential theoretical resources such as with multi-level systems modelling and complexity theory, or other socio-ecological approaches (Dooris, Poland, Kolbe, De Leeuw, McCall & Wharf-Higgins, 2007). Sometimes, though more rarely, we see a focus on the potential measurement issues that are involved in dealing with these more complex, contextual issues (Camposstrini, 2007). Seldom, is there enough emphasis on the basic foundations of conceptualization. It is here, down in the dirt so to speak, that the pillars of an alternative approach to gathering evidence of health promotion effectiveness will stand or fall. Unlike the truly great systems theorists (Luhmann, 1982, 1995, 1998; Parsons, 1951, 1977), health promotion has seldom put requisite efforts into the basic theoretical work necessary to get coherent, rigorous concepts off the ground.39 Furthermore, what will be argued below, is that even when ‘conceptual work’ is done, little attention is paid to the actual work of making empirical materials match categorical distinctions; in other words, how do we actually make coherent, defensible abstractions; how do we actually, with accountable consistency, code instances or cases just as an example of this category or that category? Yet, we know from extensive research in laboratory studies of the sciences, that workable research programs must have a solid, everyday foundation of situated, accountable, recognizable and repeatable methods for collecting, producing, filtering, identifying and literally seeing their collective materials as specific examples of ‘things’ of interest (Lynch, 1985).

39 There are some emergent exceptions to this criticism, such as in the recent collection by McQueen & Kickbusch (2007).
With realist synthesis, we have a very specific approach to conceptualization and abstraction. For realists, ‘abstraction’ is an achievement verb. When a realist has accomplished a valid abstraction, they have isolated the essential aspects of an object or a relation from the inessential aspects. It is an abstraction from the concrete, yet the move is to synthesize judicious abstractions to better understand the concrete whole. Furthermore, when identifying abstract relations of objects, we must distinguish between substantial/formal relations, internal/external and necessary/contingent relations. These are just a few of the many distinctions and wonderfully clear, concise guidelines offered by easily the most practical book on modern realist methodology and method for social science (Sayer, 1992).^40^ However, despite the clarity and the comprehensiveness; despite the easy to follow examples and the stimulating thought experiments; every new scene of this great accomplishment of abstraction is accompanied by a new piece of work.

So where do we start in thinking about how to conceptualize context? One of the key thinkers in the field reiterated the complexity of the task:

"the starting point that you mentioned is that the contexts are multiple, you know, and onion-skinned and layered and that’s the crucial difference between different incarnations of the same program might simply be the ideas involved or the individuals involved or the interactions involved or the institutions involved or the, you know, infrastructural societal levels and so it’s the same problem. There are an infinite number of contexts that you could call upon and, you’re right pragmatism rules" (personal communication).

Here there is noted and adumbrated one way of slicing up the ‘context’ into a series of system levels (individuals, institutions, societies) mixed with ‘ideas’ and ‘interactions’. One could imagine many more or less elaborate lists or models of context levels or

---

^40^ I am indebted to Andrew Sayer, not only as one of his former students, but for his example of how to present complex ideas in plain, comprehensible language, avoiding the strange obscurity in which many contemporary ‘theorists’ cloak their ideas. If I can merely approach his example, I would be pleased.
categories. Here, the same thinker says that ‘pragmatism rules’. What might be meant by this is not entirely clear, but what it definitely implies is that the realist approach does not dictate some once and for all ‘way’ of looking at context; rather, it intends to encourage a ‘common sense’ or pragmatic approach based on the particular set of phenomena one happens to be interested in when conducting a particular review. In other words, what we have is a very refined, scientific sounding injunction “go out and abstract!”, combined with a very vague set of instructions about how exactly to go about doing so.

This first sequence of interaction concerns the categorization work that is seen as a necessary part of developing some more abstract conceptualization of the relevant contextual factors that seem to have been important across the range of selected case studies and project materials the research assistants have assembled. The project started with a very high level set of abstract categories derived for the ‘Effectiveness Framework’, which were used as initial placeholders for sorting contextual materials in the first phase of extracting data from the collected cases (Social, Economic, Political, Environment, Gender, Cultural). The student research assistants (RA) have compiled three sets of relevant documentary materials to work with. First, they have two large red binders, with over 100 articles between them, organized alphabetically. This is the source data, made up of published reports and evaluations of health promotion projects or programs. Second, they have spreadsheets that they have constructed in order to extract relevant data concerning mechanisms, contexts and outcomes from the source data. Third, they created a summary text, with some preliminary analysis of the extraction work.

Figure 1 ‘The work setting’
In this first scene (see figure 1 above), the project coordinator (S) and the two RAs (L and J) are in a boardroom, with a long table, on which is set the two red binders, a video projector and a laptop computer for projecting the extraction table of interest onto a screen placed on the wall across from them. Each RA has a notepad, a pen and some other supporting documents. S is sitting directly across from the screen on the opposite wall and behind the laptop, from which he is manipulating the spreadsheet projection. To his immediate right, sitting at the corner of the table is J, oriented to the screen. L is sitting to J’s right and diagonally from S. L is leading the discussion and explaining how the RAs have been trying to classify various contextual materials into the various crude categories. The excerpt below revolves around trying to sort out where and why certain extracted data fits under specific categories. Furthermore, there is a particular interaction concerning what are
the ‘appropriate’ categories to use in characterizing the types of issues that were deemed ‘contextual’ in the material, specifically concerning the category of ‘cultural’. Below is the full sequence, followed by a closer analysis of the specific steps of reasoning involved.

Excerpt 1: Context categorization devices (this excerpt runs from 2:35-4:37 of the ‘Context’ session)

1  L:  And I put, I didn’t create any sub-headings
2  S:  Yeah,
3  L:  But for me language comes under culture, and so does gender, race, ethnicity, religion, age, education and then as a separate ones were economics, politic um politics um and history were kind of sep-separate broad categories. They weren’t as—
4  J:  You put education under culture??
5  L:  I do. Education, ethnicity… but that’s my interpretation. Culture’s a really broad concept, so um (.) you may want to…
6  S:  Yeah well that’s true. We might want to instead of having culture in that one use ethnicity maybe.
7  L:  That’s another thing, ethnicity isn’t, these aren’t all ethnic that’s the only problem. When they talked about cultural awareness it wasn’t always ethnic so that’s why I had a problem with it.
8  S:  What did they, what did they use other than ethnic.
9  L:  Um what they usually spoke about was, I don’t know, there’s a lot of links to language,
10  S:  Yes,
11  L:  Which is a separate category,
12  J:  I think under culture it was mostly ethnicity it was like dealing with Mexicans or Spanish or First Nations but…um
13  L:  But this is only some list of all of this, all of this together,
14  S:  Right,
15  L:  You know what I mean like, for race and ethnicity only two actually said,
16  S:  So what are we—are we just saying that culture’s not necessarily the most helpful category?
17  L:  No these are just ones that they didn’t really explicitly say what was culturally sensitive about it ‘cause they didn’t mention race or ethnicity they didn’t say we took participants from this background, this background, it was just the idea that these were ones that were very broad and they just said this is culturally sensitive,
18  S:  Okay,
19  J:  But they didn’t talk about it (.) and the ones that did actually mention they are the ones under language, gender, race, ethnicity and I don’t know, anyways (.) we can look at them further to make sure
The sequence begins with the acknowledgement by L in line 1 that she hasn’t created any “sub-headings”. By this she meant, she hadn’t started to systematically create more refined sub-categories for each of the high level categories of ‘gender’, ‘cultural’, ‘environment’, etc. S says ‘yeah’ in line 2, to acknowledge that this is fine and we can just move on. Of course, this is a kind of ‘report-acknowledge’ pairing that is normal for the relationship of the project coordinator (PC) and an RA. However, what L is actually initiating in lines 3-6 is really a statement/question. This is a different kind of invitation; it is no longer a report of something that hasn’t yet been done; rather, it now acts more as an invitation to consider a sort of preliminary effort L has made to think through the problem of ‘sub-categorization’, tied to a implied question (given the PC/RA relationship) as to whether this specific line of reasoning makes sense. In line 7, J interrupts L with a direct question, with rising intonation, indicating some scepticism about L’s classification of education as ‘under’ culture. This interjection of J’s is not mainly about disagreeing with L; instead it acts to signify to S that this is an issue that the RAs would like to discuss more. Very likely, both L and J have already interacted with this generic problem, if not this specific classification item. What is being done is the necessary work of getting the PC, who has not been immersed in the data at this point, to orient to what the RAs have already identified as a ‘problem’ or at least ‘issue’. While the RAs are not as familiar as the PC with the theory of realist synthesis, they are much more familiar with the actual ‘task-at-hand’, which is to accountably classify extracted materials under the ‘right’ category. What the ‘right’ category is, is precisely what requires some discussion/clarification. Here we can pay a little closer attention to lines 8-14:

8    L:    I do. Education, attainment... but that’s my interpretation. Culture’s a really broad concept, so um (.) you may want to...

9
S: Yeah well that's true. We might want to instead of having culture in that one
use ethnicity maybe.
L: That's another thing, ethnicity isn't, these aren't all ethnic that's the only
problem. When they talked about cultural awareness it wasn't always ethnic
so that's why I had a problem with it.

In line 8, L responds to J's question in the affirmative, "I do". She then moves to
justify this 'interpretation' by referring to 'culture' as a "really broad concept". S interrupts
in line 10 and moves too quickly to a 'solution' by substituting 'ethnicity' for 'culture'.
Whereas, for L, it is not just that 'culture' is a broad concept, but that there is no easy
substitute, particularly not 'ethnicity', because "when they talked about cultural awareness it
wasn't always ethnic" [my emphasis] (line 13). This raises a series of classically
epistemological questions. First, how do we classify certain statements as standing for some
particular social science category? What criteria do we use? L is suggesting that an important
criteria is that the sources of the data, which are usually the writers of reports or participants
being directly quoted, must state directly or at least strongly imply that the types of 'issues'
that they deemed important contextual factors are in fact issues of 'ethnicity' if we are to
categorize them as such. After a question from S concerning what terms were used, if not
ethnicity (line 15), L mentions language and that this is a 'separate' category (line 19). Then
J makes an important intervention, because in lines 20-21, J outlines an alternative
interpretive strategy:

J: I think under culture it was mostly ethnicity it was like dealing with Mexicans
or Spanish or First Nations but...um

Here J is saying that while the sources themselves might not actually mention
ethnicity, the fact is that they are often from specific ethnic groups, so when they mention
cultural sensitivity, we can legitimately classify this as related to ethnicity. The important
thing at this point is not to 'choose sides'. Whatever we think of the best strategy, we can
clearly see that there is an extraordinary amount of complex practical reasoning that underlies what at first glance seem to be very simple and inconsequential decisions. Yet, if we take L’s implicit epistemic reasoning as our guideline, we must be much more restrictive of our typical social science impulses for inferring from a pre-set categorical assumption what is important and relevant in a source’s report. If we were to move quickly to accept J’s reasoning as a legitimate move, it allows us to not only ‘fix’ a categorization problem, but will allow us to smuggle in all sorts of theoretical assumptions about ‘ethnicity’ into our later analyses. In lines 25-26, S moves to wonder whether ‘culture’, because it is so broad, really is going to be a ‘useful’ category. What does S mean by useful here? Remember, the purpose of the realist approach, unlike a strictly narrative analysis, is to develop mid-level abstractions concerning context factors so that a comparative logic of analysis can get off the ground. While we may not be looking for variables in the classic sense, we are (that is the research team) looking to have abstract contextual concepts and categories that will do the work of providing resources for explaining how different contexts interact with mechanisms and thus produce differences in patterns of outcomes. Thus a category becomes ‘useless’ if it applies to nearly everything. Yet, in the rest of the interaction (lines 27-31; 33-35), L strongly advocates that we don’t really have the inferential resources to make the step from the fact that people say something was about ‘cultural sensitivity’ to the conclusion that it can be slotted into a more refined, mid-level category. There is no real correct answer, other than that a decision has to be made about what to do with those pieces of data that fit as ‘other’. What sorts of inferential strategies are legitimate, consistent and therefore accountable and repeatable. Furthermore, as a matter of immediate practical necessity, the RAs need to be
able to continue their work by acting on some ‘rule’ or ‘procedure’, no matter how indeterminate is must end up being.

We move from this interaction through a series of direct investigations of the extract table in the spreadsheet. We can see from the excerpt below how an unfolding pattern of concerted work utilizes the spreadsheet tool, the binder archive and the collaborative glossing efforts of the PC and the RAs to elaborate the preceding theoretical discussion as a theme or issue that exists in and through the materials. As Lucy Suchman, a pioneer in ethnomet hodological studies of work, puts it: “[a] central concern of studies of work has been to show the mutual constitution of knowledge and artefacts in practice” (2000: p.43). One could almost say that the central ‘policy’ of ethnomet hodology is to undermine the notion of a mentalistic epistemology, where ‘knowledge’ is created and resides in the mind of the theorist and embodied and enacted practice is relegated to the realm of afterthought. In relation to ‘science’ this prejudice is, of course, a pervasive ideology. In the context of a nascent alternative methodology, there is even more of a tendency to assume that the most important work to be done is theoretical. While I have no intention of denigrating theory, part of the thrust of this study is to show how the so-called mundane, practical work of the RAs is actually a core element of the potential success of the project itself, but also represents the kind of foundations that will make possible the future development of the realist approach, and in fact, any other viable alternative to the EBM model. What we have below is what can be described as a primitive socio-technical system (primitive because this is literally the first iteration).

Excerpt 2: Working with the ‘context’ materials (this excerpt runs from 4:40-7:09 of the ‘Context’ session)
S: Okay, well let’s just have a look at one of them, let’s see, (manipulating laptop to control projection of the spreadsheet table on the screen)
J: Yeah pick one for sure
S: um (. ) let’s go up here for a sec
J: If you go up a little bit you’ll get Kiefer
S: (reading screen) Moderators and assistant moderators from the community had oral and written bi-lingual Spanish-English ability, culture was a prime factor in influencing family a::nd—
L: You have to click on it (1.5)
S: (reading screen) Community patterns of eating, okay (. ) participants described
L: See so that’s=I think that--
S: So, culture of the (. ) of the actual Latin eh (1.0) group, or what was happening there?
J: Well we can get that article out too, these are just notes,
S: Yeah,
L: This i::s, what is it?
S: Kieffer
L: Kieffer?
S: 2004
J: It’s the other one (gestures to the red binder further from L)
L: I think culture’s a bad topic with me beca::use, I come from Anthropology so::—
J: You can also look at Mccoley-Macaulay and Maciak (gestures to the screen)
L: Yeah
J: So, I don’t know, (gestures at screen with her right hand) I can’t see where the lines are crossed but what does Macaulay say?
S: (reading screen) Culture relevance and creating programs with congruency to the common values of the participants (stops reading (. ) so this is about the aboriginal Mohawk tradition.
J: Oh yeaah Mohawk just taking care of the children’s future and then the Maciak one is also (. ) under culture (1.0) (reading screen) great CBR no males, trust issues ownership, culture differences
L: Okay (finds the article)
J: (3.0)
S: That’s the Maciak there, right?
L: S:o, the one by Kieffer it only has that one-itsa same blurb, it says (reading article) moderators and assistant moderators for the south-west focus groups had oral and written bi-lingual Spanish ability (. ) a::nd (1.0) culture was seen as influencing family and community kinds of eating in particular.
Participants disagreed about whether diabetes was inevitable for those with a family history of the disease
S: Right,
But that's—that was all—but they were talking about the impact of family on their

Yeah

And to me that wasn't necessarily ethnic but

In line 36, S signals that they should now turn to the 'synthesis table' that is projected onto the screen on the far wall. They key feature of this excerpt is its orientation to 'repairing' the interpretive dilemma uncovered in Excerpt 1: the problem of how to classify different contextual 'instances' from the case studies. At dispute, or at least in question, are the categories of 'culture' vs. 'ethnicity'. The work now being done is to 'go through' the synthesis table as an initial step, and then to refer to the full article in the red binders, if necessary. This is what forms a primitive socio-technical system. S is leading the inquiry.

This is evidenced by his role as 'questioner' in lines 49-50. He is in control of the laptop and is manipulating the cursor along the rows and columns of the synthesis table. The two RAs are playing the key role of helping S to navigate around the table. The RAs have developed and frequently used this table to input data and therefore are well acquainted with its functionality. This is not true for S (this is the first time he has worked with the spreadsheet-based table). You can see some of this navigation work by J in lines 40 and 62, and by L in line 45. The red binders also form a key part of this system as they are the 'ground zero' of the inquiry environment; they contain the 'raw data' which forms, for all practical purposes, the ultimate location for the trajectory of each line of inquiry. I call this primitive socio-technical system a 'proto-objectification machine'. What is meant by such a strange locution? Well, let's step back to the problem at hand in Excerpt 2. Remember, that what is 'problematic' is the threat to objectivity that is formed by the 'subjective dissension' between L and J and S as to how to properly categorize specific examples from the case study materials. Porter's distinction between 'disciplinary objectivity' and 'mechanical objectivity'
was referred to in Chapter 1 (p. 20). This ‘proto-objectification machine’ is a kind of mundane, everyday under-labourer for either of the two types of objectivities. In fact the former types can be thought of as accomplishments, while the latter is a kind of complex process. From an ethnomethodological point of view, this type of process is not unique to scientific work; rather, it is a ubiquitous feature of everyday life. The mutual constitution and recognition of what is the case in any particular context is part of a morally guided practice, where members of social life reflexively report on, make accountable and establish states of affairs as mutually objective states of affairs. Furthermore, and crucial to Garfinkel’s work in particular, is that this ‘practice’ is made up of temporally sequenced interaction orders (Heritage, 1984: 108). In relation to this sequence of interaction, the issue is to reflexively establish the sense of the categorization process as it was applied as an accountable activity, based on a reasonable course of decision-making. In lines 8 and 14 of the previous excerpt, L brings attention to the subjective character of her reports (“but that’s my interpretation” and “that’s why I had a problem with it” respectively [my emphases]). The sequence then becomes a concerted effort to help to repair the subjective dissension. The rest of the sequence consists of a series of direct inquiries lead by S, going through specific examples where the tension between the categories of ‘culture’ and ‘ethnicity’ can be topicalized (from the table, ‘Kieffer et al. (2004)’ and ‘Macaulay et al. (2007)’). In lines 41, 47 and 66, S reads directly from the table the extracted summary of the articles. In lines 51-59, a short sub-sequence is enacted where J suggests we can look at the article (line 51), S agrees (line 52), L and S confirm which article is needed (lines 53-56), J directs L to the correct binder (57) and L gets up and retrieves the binder (58-59). This sub-sequence is a key part of the machinery
of the inquiry process, and is repeated throughout the work sessions. We will return to the
significance of this pattern of interaction later. In lines 60-61, L makes a critical intervention:

60      L:     I think culture's a bad topic with me because, I come from Anthropology
61      so:—

Here L is making a simple point, but actually quite a profound one. As a Masters
student in Anthropology, she is raising the issue of how the category of ‘culture’ is not just
simply one among others for her, but not because of her individual eccentricity; L is making
the point that a whole discipline (Anthropology) finds this category problematic (and, that is
indeed the case, as any perusal of an introductory text in Anthropology will demonstrate). As
Raymond Williams says in his book Keywords, “[culture] is one of the two or three most
complicated words in the English language” (Williams, 1976: 87). In fact, the subtitle of his
book is “A vocabulary of culture and society” [my emphasis] and in it he devotes a full six
pages to the discussion of this one word. What L is making topical is not only that ‘culture’ is
a ‘bad topic’ (which refers directly to how it is generating problems for the task at hand), but
also that she is bringing specific resources to this task and that they are disciplinary
resources. In a sense what is being said is that, even if the local team (consisting of S, L and
J) can’t come to a consensus, this doesn’t leave us with a fall back position of absolutely
atomized and isolated subjectivity; rather, there can still be some sense made of the situation
in terms of different collective understandings, here based on a disciplinary perspective.

The rest of the excerpt revolves around considering how to categorize a particular
intervention in a Aboriginal Mohawk community (Macaulay et al.) and one in a South West
US Latino community (Kieffer et al.). Here again, there is yet to be a consensus, but at this
point the team is being guided by L through her own course of reasoning about the
‘problems’ she had with the category ‘culture’, but also with her difficulty with the category
‘ethnicity’ as well. Finally, L now is the one that reads directly from the table (lines 75-80). She ends by reiterating her concern about lumping even this example under ‘ethnic’ (line 85), yet with the qualifier ‘but’, indicating that she might understand that we use that category for other overriding practical reasons. Despite the appearance of some unfinished business here, these initial two sequences, along with some other similar issues elaborated in the interactions that followed immediately afterwards, lead to the conclusion developed in the last sequence in this section, as we see below.

This excerpt is entitled ‘What’s ‘interesting”’ which comes from the sub-sequence beginning at lines 111-112, where L is asking a question, and ending at lines 118-119 where S uses the locution concerning what’s “not really that interesting to know”. This particular locution is repeated at different times throughout the transcripts (lines 131, 133, 210, 247, 252 and 279). In each case, it acts as a kind of inquiry guidance mechanism. It becomes a shorthand gloss that stands for the more complicated reasoning that lies behind the idea of ‘what sort of procedures, concepts and tools should we be using to get at the essence of the research questions we are interested in answering’. The whole set of ambiguities we find when we deconstruct the play of terms interest/interested/interesting in the sequence of interaction below highlights several key issues relevant to our more epistemological concerns. First, ‘having an interest’ in something can be thought of as the ultimate ‘subjective’ trope, connoting a conscious intentionality, and even at its extreme, naked ‘self-interest’, or in the more collective parlance of contemporary politics, ‘special interests’. It can be thought of as the very definition of being ‘partial’ or ‘biased’. However, to say something is ‘interesting’ means that it ‘causes curiosity’ or ‘holds the attention’ of an observer (OED, 1995). The latter is often the notion held up by scientists to defend
themselves, precisely when they are being accused of having illegitimate interests that contaminate their scientific work. In other words, the ultimate value and basis for impartiality in science is a subjective attitude of pure scientific curiosity. In other words, the condition of possibility for the production of ‘objective’ scientific inquiry relies ultimately on this cultivated attitude within the scientific community. The necessity of this ‘public image of the scientist’ is in place, irrespective of the often decidedly private hypocrisy that underlies many a scientific career.\footnote{41} It is also, of course, recognizably a key part of the formation of the ‘professional social ideal’ (Perkin, 1989). In the excerpt below, the team pulls back from the immediate problem of fitting examples to appropriate categories, and creates a collective space (and time) to reflect on what types of categories we actually require in order to carry out the work in the best way. What we have is a perspicuous example of what Mike Lynch calls ‘epistopics’, where we can see empirical ways in which classical epistemological themes have to be worked out in practice by members in situ, whether they be astronomers ‘discovering’ a pulsar (Garfinkel, Lynch & Livingston, 1981), jurors making deliberations (Garfinkel, 1967), or the struggle between ‘testimony’ and ‘evidence’ in the Iran-Contra hearings (Lynch & Bogen, 1999).

Excerpt 3: What’s ‘interesting’? (this excerpt runs from 32:05-33:57 of the ‘Context’ session)

86 S: So we gotta think-start thinking about how to categorize some of this, because
87 umm...
88 J: Yes, yes
89 S: you see it might, it might very well be that these=the broad categories are just
90 not very helpful
91 J: mm hmm
92 S: (.) uh especially when ehm:: (2.0) ah especially when some of the problems
93 are very specific...
94 J: They’re very yeah

Uhmm

The thing she said about uhmm (1.5) (gestures with her right hand to the
to screen), difficulty when trying to transfer skills so they were doing some
training program and because everybody’s changing

Yeah (nods head)

they would train people

Yeah (nods head)

=and they’d be gone

They’d lose the knowledge yeah

So the contextual problem there was just the fact that there was just this crazy
turnover that wasn’t facilitating good programs

Okay

=Just so you understood, sorry

Yeah

Uhmm (.) (rubs forehead and looks back to the screen) yeah:::, in terms of
categories

Should we just create, so would you just have economics (gestures with right
hand to indicate levels), and then have all the sub-categories?

Yeah::: uhm we could keep them for a little while but I’m worried that eh if we
spend too much time (.) eh worrying about fitting them into those really broad
categories that we’re missing what actually is being said in terms of the
context

Okay

Uhm::: because its not really that interesting to know everything fits under
economics or everything fits under
culture

No (shakes her head emphatically)

culture

I don’t think that’s useful really, I think its more useful to look at the
individualities of day-care

Well childcare for example for me is a really

that’s a hu—

and then we can talk about how its got cultural, gender

yeah

=and economic—

yeah

=components, but the issue is childcare

I think its more interesting to look at the actual issue of like the use of use of
words like not just language but the use of actual words like the connotation
they come with, the history that word comes with, like that’s an interesting
issue

Yeah

(.)

but if we just wrap them all up (gestures with her hands in a whirling motion)
in economic and then try to write something about how ‘economics’ effect

Yeah, no that’s not gonna be good
J: of course it does, everybody knows tha—

So, we start in lines 89-90, where S notes that the broad categories the team started with may not be "very helpful". What does this mean? To understand the problem we have to go back to the start of this section, where it was noted that there is a very clear and powerful injunction by Sayer to take 'abstraction' very seriously. Sayer begins his discussion by noting that in science, abstraction is often 'taken for granted'; in other, words, what is seldom paid much attention is "what we abstract from" [original emphasis] (Sayer, 1992: 86). Sayer says:

"...in using the metaphor of 'reproduction' for describing social processes, sociologists need to consider the costs of ignoring their open-ended nature and their dependency on skilled actors. Often the abstractions will indeed prove safe, but simply using them out of habit because they seem redolent of 'science' is hardly a recipe for rigour" (ibid.)

I use Sayer as an example because his is a very detailed account from a realist methodological perspective of what makes for judicious and appropriate abstractions in social science. He outlines some classical philosophical distinctions concerning objects and relations and mixes this with some easy to understand examples and some questions the potential investigator can ask to get their inquiry off the ground. Yet, in any particular case, what is it that makes an abstraction 'helpful', 'useful' (line 122), or 'interesting'? It is one thing to suggest that you only abstract the essential qualities of things, and make sure you don't abstract from and thereby occlude what is really important about a phenomenon. Or to say, along with Alfred North Whitehead, be sure not to commit the 'Fallacy of Misplaced Concreteness'. Here it is worth quoting Whitehead in full from his 1925 Lowell Lectures:

"...if the abstractions are well-founded, that is to say, if they do not abstract from everything important in experience, the scientific thought which confines itself to these abstractions will arrive at a variety of important truths relating to our experience of nature [my emphasis]...[the] disadvantage of exclusive attention to a group of abstractions, however well-founded, is that, by the nature of the case, you have abstracted from the remainder of things. In so far
as the excluded things are important in your experience, your modes of thought are not fitted to deal with them. You cannot think without abstraction; accordingly, it is of utmost importance to be vigilant in critically revising your *modes* of abstraction.” [original emphasis](1967: 58-59).

These are two very sound pieces of theoretical advice, nearly 70 years apart. In fact, it is hard to think of how one could give a better ‘recipe’ for conducting abstractions. What it leaves out, however, is how one applies this advice in the course of everyday scientific work. This is not a demand for a fuller description or recipe. After all, at some point every recipe must end, or what is gained in verbosity is lost in utility. Yet, what is left out is not a minor detail. To see this we can follow our example above. Of course, as noted in Whitehead’s wonderfully concise description of the issue, judicious abstraction turns on the whole trade-off between *specificity* and *generality*. Indeed, one could argue that most of the epistemological disputes in the social sciences can be traced back to this basic dilemma. We see S raising this issue directly in lines 92-93, and then J using a specific example to elaborate this theme in lines 96-105. In lines 99, 101 and 103, S encourages and agrees with this elaboration. This example/agreement pairing is a pervasive phenomenon in all of the work of the team, and it a primary ‘objectification mechanism’. We see this repeated throughout the transcripts (it is repeated below in lines 124-130, except with a role reversal between S and J). In lines 109-110, J begins to direct an open question (i.e. to the group as a whole) what the next step should be, to which L responds in lines 111-112:

111 L: Should we just create, so would you just have economics (gestures with right hand to indicate levels), and then have all the sub-categories?

112

Here L is asking whether the strategy should be a ‘top-down’ one of building ‘sub-categories’ under each main category (a strategy that had been discussed in an earlier session). S, in lines 113-116, starts to build on his initial concern about specificity, and then in lines 118-119, he further elaborates with the shorthand gloss about what’s ‘not really that
interesting’ about the top-down strategy. From 120-142, J and S initiate a turn-taking activity of starting to build an alternative strategy of working from the ‘bottom up’, with specific contextual issues that are evidently arising in many different cases, such as ‘childcare’ (lines 122-124). In fact, in the latter case, both S and J say ‘childcare’ and ‘day-care’ simultaneously and independently, when automatically searching for a good example of one of these ‘on the ground’ issues. In lines 131-134, J takes up the topic of what would be ‘more interesting’ and outlines what appears to be a loosely sketched plea for more attention to the discursive construction of the case studies themselves. In lines 138-139, J uses a key device of prospectively considering the implications for ‘writing up’ the results of the research project. This is a key cognitive strategy that is a criterion for giving some substance to the answer to the question ‘what’s interesting’. As she says in line 142, ‘everybody knows’ that economics has an effect, but precisely because of this, it can hardly be considered ‘news’ or an original conclusion; we need more specificity.

We move, in the next section to a series of excerpts that relate to theme of the epistemology and ontology of ‘outcomes’. We see that these issues are essentially intertwined and that a reconsideration of how we conceptualize this critical area is fundamental to building alternative approaches to assessing health promotion effectiveness.

The Ineluctable Ambiguity of Outcomes

“Ineluctable modality of the visible: at least that if no more, thought through my eyes. Signatures of all things I am here to read, seaspawn and seawrack, the nearing tide, that rusty boot. Snotgreen, bluesilver, rust: coloured signs. Limits of the diaphane. But he adds: in bodies. Then he was aware of them bodies before of them coloured. How? By knocking his sconce against them, sure. Go easy. -James Joyce, Ulysses

As Joyce reminds us, the ‘diaphane’ is certainly, limited; yet, whenever we go out in the world to find out what is there, we must ‘knock our sconce’ against it. As McQueen has
recently stressed (McQueen 2007b), the attempt to assess the effectiveness of health promotion struggles with a highly complicated ‘epistemology of outcomes’. I would go further and say that this difficult epistemology is also caught up in a bizarre and ambiguous ‘ontology of outcomes’. Whatever, the theoretical impasses that present themselves when working with this issue, in terms of getting the synthesis review done, it was necessary that, for all practical purposes, we made decisions about how to conceptualize outcomes. Following the ‘effectiveness framework’, we had at least five separate categories of outcomes that were deemed relevant as markers of effectiveness of health promotion interventions: individual outcomes; community outcomes; system outcomes; changes in determinants of health; and, population level health outcomes. This work follows a long course of development within evaluation (Suchman, 1967, Patton, 1997, Mayne, 2001) and health promotion (Nutbeam, 1999) that has recommended that complex social and health interventions be analysed with a ‘means-end’ framework that includes a ‘chain of outcomes’, ranging from the ‘immediate’ process or implementation objectives, the ‘intermediate’ or short-term outcomes, to the longer term ultimate objectives (often referred to simply as ‘outcomes’ tout court). There has been much work in this area and many different models of ‘theory-based’ evaluation (Lipsey & Pollard, 1989; Chen, 1989, 1990; Chen and Rossi, 1989; Patton, 1997). Some, such as Patton, offer quite practical resources for considering how to use the theory of a program to build up a picture of the key elements in the program and therefore how to go about conceptualizing the ‘means-end’ continuum. In parallel to Sayer’s more generic methodological advice, Patton cannot be faulted for his efforts. Similarly, the WHO Evaluation in Health Promotion book, or the ‘yellow book’ (WHO, 2001) (reviewed
briefly in Chapter 2), has many nuggets of sage guidance. However, once again we are left with a lot of work to do.

The RAs had been working with the ‘Resource Guide’ to the framework. This guide gave definitions of the different levels of outcomes. However, as can be seen in the sequences below, the work of identifying outcome ‘instances’ as fitting into one or other category was not so straightforward.

Excerpt 4: Moving boundaries (this excerpt runs between 2:08-4:27 of the ‘Outcomes’ session)

143  L:   Now, the way that we do it—I guess that we distinguish between individual change and community change?(. ) was for the individual change it was mostly what the participants reported, so just the participants themselves and by that I mean like people that are actually involved in the interviews and the focus groups and the training sessions(. ) what changes they reported themselves going through and then for the community change it was more what happened in the broader community (. ) for me that was just difficult because they said a lot of ‘increased this’ and ‘increased that’ but you didn’t always know if it was just that small population that they were talking about or the entire community
153  S:   Right--
154  L:   So:::
155  S:   =right
156  L:   Some of these(.) someone might have to go through these (gestures at screen) to see the difference between those, but I tried make it as much as possible discernible...
158  S:   Well-when we’re talking about the percentages, what percentage of ahh-articles just didn’t have any(.) discernible...
160  J:   That’s why there’s ‘other’ (points to the ‘summary’ paper in front of her) that ‘other’ has both, that’s 41%, those are ones we don’t know where they fit
162  L:   Yeah, but you mean nothing right?(.) we didn’t put that
163  J:   There’s none that had none(.) I told you that I was left with four articles--
164  L:   You said four yeah
165  J:   =that I couldn’t find anything for--
166  S:   anything at all?
167  J:   =and that’s after I’d done some searching, so there should be four that have-- and these are of course overlapping so some of them will have reporting of a whole bunch of different types of outcomes and some will just report one and
170  S:   okay(.) yeh
172  S:   okay (. ) great
L: umm, so I guess for individual change—
S: It’s the highest
L: =a lot-it was the highest for sure and the-eh two most popular were more-
people physically active and eating healthier foods (.) and then you get into
smoking less, receiving better health care, improvements of family
involvement, less violent activity, this was—
J: significant??
L: Yeah (.) that was in a couple of them that was kind-of, I think that in spousal
relationships
J: In father-son stuff too, I read an article about domestic abuse—
L: and substance abuse, sexual behaviour::r, more knowledge and awareness,
increased collaboration, see that’s another one I wasn’t sure of—
S: Yeah, that’s good—
L: =that should be—
S: =that’s a question mark ehm (.) depends on what- -their::r—how-their…
L: Soo, that might possibly go under community change??
S: Yeah, yeah I think your right—
J: But if an individual is referring to (.) increased collabora-collaborat-well-I
guess that they have to be collaborarting with someone—
S: the change is collaborative yeah—
L: okay, so maybe what that would be—
S: I think that would probably be moved—
L: okay—

In lines 143-152, L outlines a classic issue for health promotion evaluation: how do we distinguish between impact on individuals that participate directly in a program and more ‘community-wide’ impacts? Sometimes this issue is quite straightforward; however, as L notes, depending on how results are reported this distinction can be very murky. In lines 149-152, L complains that sometimes the ‘report’ does not make it clear if its impact on individual change was restricted to direct program participants, or did it show up in the wider community. As L says in lines 156-157, it is likely that we will need to go back to the source data to clarify this issue. In 158-159, S switches the focus to an overall question concerning how many of the 101 cases had no information on outcomes. J answers (lines 160-161) that those are under other, which make up 41% of the cases. J understood the question to be about those cases that did not have enough information to classify them under one of the existing
categories. In line 162, L interjects that that was not what S was looking for. She clarifies that S wanted to know about how many cases had *no* information on outcomes at all. J replies that only four cases had no information, and in line 167, she refers to the fact that this was after ‘some searching’ (the project initiated a policy of treating the ‘case’ rather than just the published article: this meant that, even if the article didn’t have outcome information, the RAs would search other sources related to that project or case, including other documents, websites, and even direct contacts with project officers). In lines 168-170, S notes that there is a phenomenon of ‘overlapping’, where some cases have multiple reported outcomes, whereas some may have only one. In lines 173-178, L moves to summarize the findings around individual outcomes in terms of the types of outcomes that ‘showed’ up most often. S says, in line 174, that it (individual change) is “the highest”. He means here that what are classified as individual changes show up in the most cases, compared to other categories of change. In 178-182, L and J topicalize a particular outcome: ‘less violent activity’. J raises this issue in line 179 by noting with a question emphasized by rising intonation, the unusual nature of this outcome. L responds by giving more specific information (lines 180-181) and J follows by elaborating with her own example in line 182. Again we can see these small pieces of the machinery of building up, through *in situ*, temporally organized activities, objective accounts of what, for all practical purposes, *is* ‘significant’. These mini socio-technical machines are not in any way replacements for the traditional methodological accounts of how objectivity is established; rather, they point to the inevitably incomplete, and even obfuscatory nature of such accounts. Even in the most restrictive and prescriptive methodological social science strategies, such as with social science surveys, the actual work of ‘objectivizing’ is never contained in the methodological prescriptions alone (Maynard and
It is this ‘mundane’ objectivizing work that sustains and ultimately makes sense of social science endeavours as accountable, repeatable activities that can produce knowledge for public consumption and retain the requisite credibility that such endeavours must have if they are to continue to be supported and used by the wider society.

In the final part of this excerpt we see some ‘boundary work’ being enacted. In line 184, L mentions increased ‘collaboration’ as fitting under the ‘individual’ outcomes category. L notes that this is “another one she wasn’t sure of”, raising the issue for discussion. This is the kind of work that demonstrates L’s competence to perceive and understand her role as bringing to attention ‘difficult cases’. In fact, much of the learning and progress in the project as a whole depends critically on this type of simple, mundane competence. There are many incentives for younger researchers and RAs to avoid addressing difficult issues, particularly if one is afraid that such question raising activities might expose ignorance or otherwise lower the opinion their more senior colleagues or supervisors might have of them. In 187, S notes that this one (collaboration) is a question mark and through the interactive work of lines 188-195, it is agreed that it should be moved under ‘community change’.

In this excerpt below, there is a discussion of ‘systems change’. The definition of systems change that the RAs were working from comes from the Resource Guide:

“This category is also aimed at collecting data beyond individual behaviour, but instead of capturing the informal norms, practices and social infrastructure, it aims at formal changes that transform local policies, laws, regulations, funding rules and institutions. Systems, in this sense, to put it simply, are about politics, law and money (17).”

This definition contrasts with the definition of community change, which is focused on ‘informal norms, practices and social infrastructure’. This contrast is somewhat parasitic on the distinction Jurgen Habermas develops between ‘system’ and ‘lifeworld’ in his two
volume, magnus opus, *The Theory of Communicative Action* (Habermas, 1984, 1987). For Habermas, this distinction is central to understand complex, differentiated modern societies, and the retention of the concept of ‘lifeworld’ secures his foundational emphasis on communicative interaction as a space outside of the totalizing systems of law, politics and money. Nevertheless, in this instance, it functions more as an heuristic device for an initial sorting of types of potential impacts that are possible for community health promotion interventions to have. Based on previous work (Hills, O’Neill, Carroll & MacDonald, 2004), we had established that in most of the literature on the evaluation of community health promotion interventions, there was very little attention paid to formal reporting of anything other than individual outcomes. Thus, the interaction below starts with this context in mind.

Excerpt 5: Identifying ‘outcomes’ (this excerpt runs between 8:38-11:00 of the ‘Outcomes’ session)

198  L: and then, the systems change was fairly easy, for me—
199  S: yeah—
200  L: =just anything policy related
201  S: and, what do we got?—20%!—
202  L: yeah
203  S: =20% of them that reported on systems change—
204  J: That’s pretty (.). eh—
205  L: you don’t think that’s a lot?, or you do think that’s a lot?
206  J: that’s a lot—
207  S: that’s quite ehm, quite significant
208  J: =significant
209  L: a::nd ah
210  S: Now how did they uhm, now this is really interesting (.). how do they report on this, do they do it in a kind of formal, quantitative way or do they mainly, or is it mainly mention::ed that something changed or like how was it reported do you remember?
211  J: yeah uhhmm—
212  L: I don’t think it was quantitative but—
213  J: No, I think it was more mentioned in like re-report of findings where they would say umm like the Ministry of Environment you kno::w attended one of the meetings or something and this was an outcome (.). uhhm—
214  S: Now would they uh-uhhmm, there’s another issue—would they actually call it an
‘outcome’, like would they—would they have like a category where—
J: yeah, yeah
S: --they say ‘results’—
L: some did—
S: or ‘outcomes’
L: some were very easy—
S: where they would report that under that?
J: =yeah, and especially when I: we can go to one of them (points to screen) that has systems change and I can see If I can find an example(.)—
S: sure—
J: if you’re interested
S: I think that’s a good idea (. ) let’s see what we’ve got here
L: most of the ones where we have outcomes would have put policies
J: I think its left (gestures to the screen with right hand)
S: Left?
J: yeah
S: well, let’s just plunk in here (manipulates the laptop keyboard and scrollpad to navigate the on-screen spreadsheet table)
(2.5)
J: umm (looking at screen) individual change, community change, I guess its right, sorry
L: yeah (. ) see this one stops (gestures at screen)
S: there we go
J: oka::y
L: its really messed up, that’s why you were confused
(2.0) (two names from the table are mumbled by L and S as they scan the table)
J: That’s an interesting one that ‘Jason’—
S: out- (reading from screen) this collaborative effort was that the police officer who was associated this change effort ultimately became a national leader who was able to influence policy at the federal (. ) level-
J: hmm
S: see now that’s interesting because I think that actually, on the model (. ) is both (gestures to the screen) uhm some sort of change (. ) but it also fits within the mechanism of—
J: recruitment of—
S: recruitment of new champions, right (. ) which is, for me, I tend not(.) an again I’ve always talked about it, to me, the outcomes depend on ah (. ) how you perceive what you’re trying to do—
J: mm-hmm
S: you know like, so you could look at a mechanism and say well what we want to do is increase critical dialogue—
J: right
S: and call that an outcome, it’s just normally with these programs they would never be able to say that’s an outcome—
J: right, funding may not (laughs) continue
We can see S’s surprise in lines 201 and 203 that a full 20% of the cases reported in some way on ‘systems changes’. In lines 205-207, an interactive sequence takes place that establishes the sense of this. First, L wants to confirm in line 205 what sense to attach to S’s ‘surprised’ reaction. Was it because it is too low or too high? In lines 206-208, we find more mutual reinforcement devices where J and S confirm their interpretation of the proportion as ‘significant’. However, partly because of the surprising number of instances, S initiates a sceptical inquiry in lines 210-213 that raises a key ‘epistopic’: how do we know this? How was it reported? Was it formal or just ‘mentioned’? In lines 214-218, L and J start to recollect that it was probably not ‘quantitative’ or very formal. In lines 219-220, S inquires whether the reports themselves ‘actually call it an outcome’. This is a key question that goes to the heart of the whole extraction enterprise, because if it isn’t ‘actually called’ an outcome, then that means the RAs had to ‘infer’ that what is being described as a change is actually an ‘outcome’, even though the authors of the report themselves don’t necessarily see it that way. Between lines 221-226, J and L confirm for S, from memory, that indeed, these instances were reported as intentional results or outcomes of the projects. In line 228, J suggests we look at an example. This is a pattern that gets continually repeated in two to three steps: first, the RAs respond to inquiries from S by immediate recollection. It should be noted that this is not based on simply individual memory, but is a sort of collective remembering (Middleton and Edwards, 1990; Shotter, 1990), that is constructed through interactive sequences such as the link between L’s statement at line 225 “some were very easy” and J’s adjacent “yeah” statement at line 227; second, there is the step of referring to the ‘examples’ in the spreadsheet table (as with J in line 227). This step is almost always taken as an immediate confirmation device of what the RAs have remembered, or as a support for the failure of
recolletion. The table is the first place where a technical interface becomes part of the inquiry system; and third, there is the final step of referring to the original source article or report. These three steps are interweaved recursively throughout the work of the RAs in order to bring ‘evidence’ to bear on a series of inquiry topics. Step three is not always taken, if it is thought sufficient to refer only to the extracted data in the table.

In lines 233-246, L and J help S navigate the spreadsheet table to find a perspicuous example. In line 244 L refers to the table being ‘really messed up’. In fact, the table was not as bad as that sounds; however, it shows L’s sensitivity to the nature and temporal flow of the inquiry process. It is important that the technical tools be calibrated to facilitate the organization of the inquiry process, so as the creator of the table, L is sensitive to its inadequacies in this initial iteration or version. This seemingly unimportant element in the interaction is actually critical. Much of the success or failure of the whole project hinges upon how well the various documentary and representational devices can be wedded to the needs of this socio-technical system. Yet, this is a dialectical process that begins in many ways in a dependent position in relation to the ‘table’ and it can be thought of (at least metaphorically) as participating in the organization of the inquiry itself.

In line 247, J finds an ‘interesting’ example. In lines 248-250, S reads out the relevant extracted caption, which ends up raising another ‘interesting’ issue (lines 252-258) about the ontological status of outcomes within the Effectiveness Framework. The example given in ‘Jason’ could either be considered an outcome or the proper functioning of the ‘mechanism’ of attracting new champions. As S notes in 257-258, what counts as an outcomes depends very much on how you perceive the purpose of a project. However, as J rightly points out in
265, this is a severely restrained ‘perceptual environment’, as many, if not most, of these projects have their ‘outcome(s)’ defined for them by their funders.

Finally, the last excerpt switches back to the epistemology of outcomes. How do we know that the project contributed to the identified change? This is the classic issue that motivates the entire apparatus of EBM and its totemistic method, the RCT. It also links to the whole notion of ‘causality’ that is embedded in the philosophical dispute between realists and empiricist-positivists.

In the excerpt below, the inquiry moves to an examination of a specific example of ‘systems’ change that has been attributed to a project that worked on tobacco control as part of a Healthy Communities initiative. We see in this sequence the working through of the status of the claim that the project has made a specific difference by way of a system-level outcome (a new bill restricting indoor smoking).

Excerpt 6: How do we know they made the change? (this excerpt runs between 16:09-18:15 of the ‘Outcomes’ session)

266  S: so now lets now have a look and see what else we have in that category
267  J: mm-hmm
268  S: (reading from screen) new bill restricting work smoking in public places—
269  L: a lot of them were smoking related—
270  S: Now (points at screen) if we pull out—
271  J: Yeah (puts away previous article in the red binder directly in front of her)
272  S: =Minkler et al. 2006—(.) (pointing at screen) can we just-just figure out what
273  they’re doing there—
274  J: Do you want that? (starts to get up)
275  S: yeah like what are they (.) and how are they linking it to their program
276  J: (walks around the table over to where the second red binder is) yeah cause
277  thats a big policy (laughs)
278  S: well no, I mean you know what I’m saying its like—
279  J: no, it is-its interesting that they ah (leafs through the binder looking for the
280  article)
281  S: is it just ah but off the way and oh gee wasn’t it great that a new bill was
282  enacted or they actually saying (.) we did specific things to try to get a new
283  bill enacted
J: (finds the article) It was, got it here (walks back to her seat)
(5.0)
J: Indiana (1.0) 2003-2005 (.). It was for like a healthy ah households or
something—
L: what was that (points with her pencil to the underside of the article)
J: Healthy communities (2.5) (reading article)—
L: Smo—(.oh Smo—
J: yeah, it says ah (reading article) the passage and implementation of an indoor
smoking measure was the first concrete outcome to which many studies-study
participants pointed (1.5) more recently getting permission to build, raising
funds for and creating a new—a new playground on public land was described
by a member as one keynote effort—
S: so, basically what they're saying is that the members of the—
J: identified those—
S: the members of the project, the participants identified those changes as a
result
J: mm-hmm
S: at least partially, of their efforts, right
J: says (reading article) a variety of policy steps and activities were employed to
help achieve policy related objectives
S: okay
J: so, it was a goal
S: what does it, does it say anything about any specific ones?
J: (reading) in their early efforts to get a smoking measure passed for example
the partners worked to create community awareness, studied what other
communities had done and mobilized a strong turnout at the city council
meeting where they presented the survey data that summarized the problem
and proposed the bill
S: perfect
J: Yes, really

In lines 272-273 and 275, S initiates an inquiry concerning the status of this claim. He
wants to know just what is going on. In lines 278 and 281-283, he elaborates: are they just
mentioning it as a sort of contiguous event to the work of the project, or have they actually
made a link between specific, planned project activities and the outcome itself. In order to
establish this, J refers to the source article (Minkler et al., 2006): ‘Do you want that?’ in line
274. J takes a few seconds to refresh her memory and to locate the relevant passage in the
text (line 285). L is helping J by reading the underside of the article to see if she can speed up
the process of locating the information the team needs (lines 288 and 290). In lines 291-293,
J restates, in full, the passage where the outcome 'claim' is made. Here it is at least established that the project did explicitly see this change (the bill) as an outcome of the project. In lines 296-301 S and J collaborate on confirming this fact. In the next sequence (lines 302-305), S and J establish that this was an explicit goal (line 305) and that the project took specific policy steps in order to reach that goal (lines 302-303). The final level of epistemic inquiry pushes beyond the assertion that the project took steps, to a question about whether they give more specific details (line 306). In lines 307-311, J gives directly quoted evidence that the project took a series of concrete actions related to achieving the policy objective. In 312 and 313, S and then J emphasize that this is exactly the type of reasoning that is being looked for to support a relatively strong claim of impact. Later, this will be termed an 'exemplar'.

Summary

What this Chapter shows is that a fairly elaborate mechanism of inquiry is in operation, without any specified formal rules for how exactly to go about conducting it. The point is that a more formalized and restrictive set of instructions and methodological protocols may be helpful, and in future iterations of this particular approach (assuming it has any legs), this is likely to be an important step. However, what is apparent is the sheer obduracy and depth of the epistemic work that takes place, despite the lack of more detailed protocols. It is this mundane, yet ubiquitous structure of practical activity that underpins the sense and accountability of the work of the review process. Furthermore, this work is literally irreplaceable. There is no 'time out', as Garfinkel is wont to say, from the reflexive accountability of action (1967). No matter how detailed a methodological protocol, it must always rest on this incarnate, immortal, essentially 'terrestrial' foundation. In fact, what is
adumbrated above is an attempt to give an adequate description of a "lived course of instructed action" (Garfinkel, 2007). The instructions consisted of the variety of protocol materials, methodological papers and direct instruction and training given to the RAs over the course of two months. The lived course of the activities took place mainly outside of the work sessions themselves (where some of what took place was merely reported on); yet, what has been focused on in this chapter is how, starting with a variety of materials and instructions, the team (S, L and J) worked through a temporally structured course of action that was reflexively accountable, moment by moment, as a mutually recognized attempt to make sense of the research tasks at hand. This is not some feat to be self-congratulatory about: it could not be any other way. Reflexivity is not (or at least, does not have to be), despite some authors’ insistence, some kind of special accomplishment of self-analysis and self-confession (Finlay and Gough, 2003). Rather, as Lynch says:

"... 'reflexivity' is not an epistemological, moral or political virtue. It is an unavoidable feature of the way actions (including actions performed, and expressions written, by academic researchers) are performed, made sense of and incorporated into social settings. In this sense of the word, it is impossible to be unreflexive" (Lynch, 2000: 26-27).

It is the essential nature of the 'reflexivity of accounts' that creates and sustains the very objectivity or 'proto-objectivity' on which any more refined notion of objectivity is parasitic.

The implications of this analysis are difficult to unpack; nevertheless, a key contribution is a much deeper understanding of the massively pervasive, everyday, practical infrastructure that must underpin any substantial scientific effort to develop an alternative approach to assessing the effectiveness of health promotion. This study acts as a sort of corrective to the overwhelming attention given to the necessity of developing highly abstract theory, often at the expense of paying attention to the 'practice' of doing large-scale
scientific work. It must be very clear, that to pretend that a replacement methodology for EBM can get off the ground without gargantuan practical, collective efforts to develop tools, train students and generate the social substrate of a genuine research program, is to be naïve. Only now are we becoming aware of the extraordinary amount of practical work that goes into making the technology of an RCT work (Timmermans & Berg, 2003; Daly, 2005), and also the diversity of applications and context-based and contingent methodological troubles that have to, each and every time, be overcome. Something like ‘realist synthesis’ is only, and at this point, can only be, a projective fantasy. However, like all projective fantasies, it has its function. The answer to the question, is realist synthesis a viable alternative methodology to EBM for assessing health promotion, is that this cannot be a sensible question, at least not at this time. The real question is whether more people will begin to share this fantasy, and if they share it, will they invest energy in making it come to life. Whatever the case, we can see how, whatever approach is taken to build an alternative, ‘scientific’ methodology, its foundations rely on an unacknowledged infrastructure of common sense reasoning, the very stage on which the whole drama of metamorphosis takes place; where the lived world of objective/intersubjectivity is reflexively organized and endlessly ‘validated’, ‘replicated’ and ‘generalized’ as part of the normal course of events.
CHAPTER 5
CONCLUSION AND RECOMMENDATIONS

There are a variety of conclusions to be drawn from this study. I will discuss them in three short sections: 1) general impressions concerning the state of the art in health promotion effectiveness; 2) theoretical conclusions and recommendations for future theoretical work; and 3) implications for building the ‘practice’ of alternative approaches to synthesizing evidence for the effectiveness of health promotion interventions.

Health Promotion Effectiveness

As is evidenced in Global Perspectives on Health Promotion Effectiveness (McQueen & Jones, 2007), there is quite a vibrant core of researchers, from across the world, who are actively engaged in work on developing the state of the art in the methodologies and tools necessary to advance work in health promotion effectiveness. As Dr. Rootman notes in his forward to the book, due to outside pressures on the field, it is likely that this work will continue to grow in the years ahead (Rootman, 2007). Yet, there are two substantial challenges for the field if it is to progress in this area. First, it must, as argued in Chapter 1, face squarely up to the legacy of epidemiological dominance as the guiding discipline for public health science. This is not in any way to denigrate the enormous contribution of epidemiology, nor to deny that it has a crucial role to play, not only in public health as a whole, but in health promotion as well. It is more to insist that health promotion reflect deeply on the origins and fundamental limitations of epidemiology as the science base for health promotion tout court. In particular, health promotion should pay more attention to the underlying, constitutive utilitarianism that drives the Weltanschauung of epidemiology and therefore, much of public health. Again, this is not to dismiss entirely the importance of
utilitarian thinking, especially its reminder of the need for consequentialist reasoning about the implication of our actions at a population level. Yet, health promotion should be more actively engaged in the debates concerning equality, equity and human rights that are well advanced in the fields of moral and political philosophy (Kymlicka, 1990; Taylor, 2005). We should understand better what is at stake when we prioritize fundamental values, such as ‘equity’. There has been some recent work in this area that begins to address this issue more directly (Potvin, Mantoura & Ridde, 2007). In order to transform our thinking about effectiveness into models that fully integrate notions of equity and social justice, we need to pay close attention to debates around distributive justice, as well as developing new tools that consistently integrate these values into our methods and methodologies.

Finally, we need to be clear about the overwhelming social and political forces that form the vast, often inhospitable ocean on which the modest fleet of health promotion efforts must navigate. Thinking more carefully about the lessons we can find in the discipline of political economy, along with socio-historical writing of people like Ian Hacking and Theodore Porter, will allow health promoters to be less naïve about potential progress and more circumspect about the possibilities and the place in which the field finds its home. Health promotion is a diverse field of actors, organizations and institutions and, as Kickbusch reminds us, ideas (2007). It ranges from radical social and community activists right up to some very high-level bureaucrats; more recently, major private sector actors and institutions have entered the field. Health promotion is often more comfortable with the quiet mantle of bureaucratic peace, soothed in the cadences of earnest health diplomacy. However, health promotion is essentially a political animal and thus health promotion cannot avoid the divisive and agonistic struggles that go along with politics.
We have yet to have enough discussion and debate about strategy and tactics when it comes to the place of ‘evidence’ and ‘effectiveness’ in the health promotion conversation. If the above argument is cogent, then we have to be very clear about what we are trying to accomplish in ‘playing the evidence game’ if we are not to meander into political cul-de-sacs of our own making. One large issue we need to confront is whether avoiding rigorous quantification and measurement in health promotion effectiveness is a possible strategy, given the social power of quantification in our contemporary Zeitgeist? Health promotion, if it is to ever attain the coherence of a ‘movement’ (assuming it should become one that is), needs to find a way to push forward on many fronts, while retaining an overall strategic vision that can fruitfully guide key tactical objectives. Health promotion effectiveness is simply one element in this overall vision and strategic line of action; how it should fit, or if it should be there at all forms part of our continuing debate in the field.

Theoretical Conclusions and Recommendations

In this section, I move from a more global perspective on health promotion as a whole, to some more specific, mid-range theoretical issues that affect directly the progress of assessing health promotion effectiveness. I have already mentioned the issue of equity above. I will not say too much more about this issue here, other than to reiterate the necessity for health promotion to continue to try and keep the implications of a commitment to equity in health at the very forefront of all of our theoretical and methodological work. This will require the efforts mentioned above, along with some very specific mid-range theoretical and methodological innovations along the lines suggested by Potvin et al. (2007). While equity is certainly of paramount concern, there are other key challenges for health promotion, which I will address under three headings: conceptualization, measurement and analysis.
Conceptualization is an area in health promotion effectiveness that, in my humble opinion, needs a lot of work. A huge difficulty has been the burden of what C. Wright Mills called ‘abstracted empiricism’ (1959). With some honourable exceptions, there has been too much emphasis on the rushed operationalization of key concepts, unguided by theory and not enough fundamental work that integrates and brings consistency and coherence across key concepts such as ‘participation’, ‘equity’ and ‘empowerment’. In the work on the effectiveness projects that form the basis of this dissertation, it became abundantly clear that the level of conceptualization of these key notions is underdeveloped. Furthermore, how we are to approach thinking about the ‘system(s)’ within which these health promotion activities are embedded is a question that has only just begun to be answered (McQueen & Kickbusch, 2007). There is still much work needed on explicating in detail the fundamental dimensions of health promotion, whether they be considered as mechanisms, key processes, basic characteristics, etc. This work has started, and in some cases has been going for some time; yet, in order to develop more consensus about key concepts, much more basic theoretical work is required. Of course, as health promotion nearly perfectly overlaps with sociology in terms of its concern with the entire social field, it will unfortunately share some of that discipline’s unending conceptual struggles and interminable debates. Nevertheless, greater clarity and less murkiness concerning health promotion’s key concepts would be salutary.

Measurement is an idea anathema to many in the health promotion field who are dedicated to qualitative research approaches. However, even if one ends up rejecting the push to develop measurement as the most important objective in methodological development for the field, the types of problems that valid measurement raises act as a propaedeutic to more refined conceptualization. The one thing a focus on measurement can’t tolerate is sloppy
thinking. From the realist perspective, the first objective is judicious abstraction and concept
development along these lines. Often, an enormous amount of abstract theorizing is done
under the auspices of realism before any serious attempt is made to ‘cash out’ all this refined
theoretical work with dedicated empirical investigations. This has lead some to find realism
an extremely exciting idea, only to be bitterly disappointed when they try to apply it to real
world empirical problems. This is analogous to many who have attempted to ‘cash out’
Gidden’s theory of structuration, only to find that there is little guidance or handholding
when it comes to implementing a real world empirical study (Storper, 1985). From this study
we can conclude that, not surprisingly, realism does not tell us just how to go about making
judicious abstractions, nor does it have much in the way of practical methodological
guidance for on the ground research.\textsuperscript{42} This leads me to believe that more collaboration is
necessary between classically trained methodologists concerned with making valid social
science measurements and those trying to implement realist synthesis approaches. This must
seem a strange conclusion and recommendation; however, I believe that where realism often
suffers is in its rarefied approach to the messy intricacies of bringing research onto the
ground and following through on its ideas. I hold little promise that we are likely to solve any
of the great dilemmas at the heart of social science measurement, but I am respectful enough
of the seriousness of the endeavour, that it presents perhaps the best opportunity for further
learning and in that small way, progress. It is unfortunately very rare that we are lucky
enough to have reflective and critical quantitatively trained methodologists focus on a field
that is often hostile to such approaches. We need to move pragmatically to further develop

\textsuperscript{42} Realists such as Sayer would say that this is only to be expected as realism is actually a position within the
philosophy of science, not a fully worked out social science methodology; it can only give guidance in this
realm, and often only gestural guidance at that.
health promotion indicators, picking up on some of the work already done, but with a new emphasis on linking this development to theoretical developments. From this we can learn much from the positivist program. Whatever its magnificent failures, its core understanding that there cannot be a disjuncture between theory, data and methods is important to keep in mind. The best theorists should be working with best empirical methodologists. Health promotion, as a essentially practical field of action, cannot afford the chasm that has divided theoretical development from empirical inquiry in sociology and some other of the social sciences.

Finally, we should be promiscuously open to the many potential analytical avenues that are possible hand maidens to a brighter future for assessing health promotion effectiveness. In McQueen and Jones (2007) we see many of these avenues; in addition, we have the more high level theoretical perspectives offered in McQueen and Kickbusch (2007). I have only two brief comments to make here regarding these opportunities. First, in regard to the emergent emphasis on systems theory and ‘ecological approaches’ (Green, Richard & Potvin, 1996), especially in relation to organizational development and its system-environment analyses. While it can hardly be argued that such an avenue is a huge advance on mechanical, linear models of how to conceptualize health promotion interventions, there are a few caveats. First, ‘systems’ theory has been around for a long time (Spencer, 1898; Malinkowski, 1944; Radcliffe-Brown, 1952; Parsons, 1951; Blau, 1968; Elliot, 1980; Donaldson, 1985; Luhmann, 1995). There are all sorts of different approaches to analyzing social systems. The important thing to keep in mind is that we can learn from the dead ends of past attempts to get this program of the ground. My recommendation to health promoters is to very carefully examine the history of this approach and to consider the trenchant
critiques that have been launched against it as a way of fully explaining the social world (Young, 1988; Clegg, 1990). As Young argues, these social science approaches to ‘ecology’ are parasitic on the discipline of biology; however, the use made of the trappings of population ecology is mainly metaphorical. When we move to precise conceptualization, we are often left with merely gestural accounts of how organizations and social environments are ‘like’ ecological systems. Yet, the power of the host disciplines explanatory matrix relies on highly specified concepts, such ‘niche’, ‘species’, ‘fitness’, etc. Human organizations are not isomorphic to ‘species’ or even members of species. We must keep in mind the distinction between a metaphorical-conceptual device that allows us to have a desperately needed ‘switch’ of perspective and a practical theoretical framework for guiding a concrete research program. For the former purpose, the ‘ecological approach’ within health promotion has already done its work; for the latter purpose it is not so clear why this new application of some relatively old ideas should avoid all the previous criticisms that have been directed their way.

Second, we have the suggestion that we start to look to ‘complexity theory’ as a way of analysing the complexities of health promotion settings (Dooris et al., 2007). Again, this may be a very promising avenue of inquiry; however, it seems to me a bit premature for a field that has barely figured out its core concepts and how to measure them, to move to the type of grand analytical syntheses that complexity science offers. It would do health promoters well to go back to the original work of the Santa Fe Institute around the application of complexity science to economics, to see what level of development is necessary before any real results can come from this perspective (Anderson, Arrow & Pines, 1988; Arthur, 1990; Kauffman, 1992, Bowles & Gintis, 2007). There is some promise in the
analysis of complex, self-organizing social network structures that could offer a better set of tools for understanding how community structures sustain themselves (Valverde & Solé, 2006). These highly refined quantitative analyses, however, rely on solid quantitative data to work from. We cannot yet be secure that much of what ‘matters’ to health promoters has been put on such a solid measurement footing.

Finally, I conclude with a general recommendation that whatever fruitful theoretical and analytical paths health promotion effectiveness takes, it should not neglect the overwhelming need for the development of practical skills and research programs that test, iteratively, the potential for some of these exciting theoretical perspectives, of which the realist synthesis experiment is only one. Just as McQueen recommends (2007) that we need thousands more actual health promotion interventions, we need at least tens and maybe hundreds more actual research projects to test new ways of analysing these interventions.
REFERENCES


274-294.


Springer Publications.


Parsons, Talcott (1951). *The social system.* Free Press.


