# Chapter 2
# "Explanation," Philosophy and Theory in Health Inequalities Research: Towards a Critical Realist Approach

**James R. Dunn**

## Contents



**Abstract** The relative lack of attention to philosophy and theory in research on social inequalities in health and its consequences for explanation and methodology has been described in the social epidemiologic literature. Nevertheless, the field of social epidemiology, dominated as it is by the trappings of positivism, is arguably still in need of further development of epistemological frameworks that can adequately incorporate richer explanations of the phenomena we study. Some have advocated for the adoption of a critical realist perspective for the study of social inequalities in health in order to overcome some of the difficulties described above. This chapter uses critical realism as a means to identify some of the epistemological and methodological difficulties inherent in attempts to explain health inequalities (especially the connection between social and biological mechanisms) and offer an "affirmative" approach to health inequalities research based on the insights of critical realist philosophy. In both the physical and social sciences, concerns about

J.R. Dunn (✉)
Department of Health, Aging and Society, McMaster University, Kenneth Taylor Hall 226, 1280 Main Street West, Hamilton, ON L8S 4L8, Canada
e-mail: jim.dunn@mcmaster.ca

P. O'Campo and J.R. Dunn (eds.), *Rethinking Social Epidemiology: Towards a Science of Change*, DOI 10.1007/978-94-007-2138-8_2,

epistemology and philosophy are routinely left unexamined. This chapter argues that a more explicit engagement with such questions is an important next step in social epidemiology and health inequalities research. The chapter begins with a description of the explanatory problem faced by research on social inequalities in health. This explanatory task is difficult, but it is precisely that difficulty which cannot be adequately accommodated by the unacknowledged epistemological underpinnings of social epidemiology. The chapter highlights several key features of the critical realist approach and demonstrates how these can be usefully applied to health inequalities research to solve persistent problems of explanation and methodological pluralism. Specifically, the chapter argues that the complexity of the social phenomena implicated in this research demands a more detailed consideration of what is meant by causation, "explanation" and "theory," and of what the relationship of these concepts should be to empirical research.

## 2.1 Introduction

The relative lack of attention to philosophy and theory in research on social inequalities in health and its consequences for explanation and methodology have been the subject of several articles in the past several years (Forbes and Wainwright 2001; Wainwright and Forbes 2000; Krieger 1994, 1999, 2001; Muntaner 2004). The attention to theory, philosophy and explanation in such writings, however, has been dwarfed by the explosive growth of empirical research, which, it could be argued, has repeatedly generated highly similar findings regarding the differential distribution of health status by socioeconomic status, but produced far less "explanation." What little explanation has been produced has leaned heavily towards biological reductionism (Taylor et al. 1997), while relatively little has been accomplished in filling an obvious vacuum of *social* explanation.[1] This state of affairs, in turn, has prompted calls for more theory (Krieger 2001), but it is clear that the field of social epidemiology, dominated as it is by the trappings of positivism, is still in need of further development of epistemological frameworks that can adequately incorporate richer explanations of the phenomena we study.

Arguments similar to these are made very effectively by Forbes and Wainwright (2001) and Wainwright and Forbes (2000). They advocate the adoption of a critical realist perspective for the study of social inequalities in health in order to overcome some of the difficulties described above. I share Wainwright and Forbes' enthusiasm

[1] There are, of course, some examples of works that have been at least partly successful at mounting social explanations, including the work of Kawachi and others on social capital (Kawachi et al. 1997, 2004; Kawachi 1999), Wilkinson (1996) with regards to social capital and other psychosocial factors, Hertzman and Marmot's (1996) explanations of the health consequences of rapid economic change in central and eastern Europe, and the work of a number of authors offering sociopolitical explanations (e.g., Navarro et al. 2006; Bambra et al. 2009).

for the realist approach in general, but emphasize different components of that approach. Also, rather than using critical realism just to critique health inequalities research, as Forbes and Wainwright do, in this paper I also use it to identify some of the epistemological and methodological difficulties inherent in attempts to explain health inequalities (especially the connection between social and biological mechanisms), and I offer what I call an "affirmative" approach to health inequalities research based on the insights of critical realist philosophy. While I acknowledge the important role that philosophical approaches such as critical realism have to play in critiquing existing research, I believe that critical realism has more to offer the field than simply a critique. It can be a tool for mounting explanations of complex phenomena, drawing on knowledge from diverse disciplines, not unlike the idea of critical epidemiology described in Chap. 1. In both the physical and the social sciences, concerns about epistemology and philosophy are routinely left unexamined, but I argue that a more explicit engagement with such questions is an important next step in social epidemiology and health inequalities research.

I begin the paper by delimiting what I believe to be the explanatory problem faced by research on social inequalities in health. In so doing, I underscore the difficulty of the explanatory task, which a growing number of scholars have chosen to tackle. But it is precisely that difficulty which cannot be adequately accommodated by the unacknowledged epistemological underpinnings of social epidemiology. I then highlight several key features of the critical realist approach and demonstrate how these can be usefully applied to health inequalities research to solve persistent problems of explanation and methodological pluralism. Specifically, I argue that the complexity of the social phenomena implicated in this research demands a more detailed consideration of what is meant by causation, "explanation" and "theory," and of what the relationship of these concepts should be to empirical research. The discussion that follows is guided by philosophical investigations within scientific realism (Bhaskar 1975, 1979a, b, 1989; Pawson 1989), with particular emphasis on the work of Andrew Sayer, who provides an excellent translation of realist philosophy into practical terms for use in empirical social research (Sayer 1992, 1993, 1997).

## 2.2 The Challenge

Stated briefly, I take one of the main challenges of social epidemiology, and the focus of the remainder of this chapter, to be related to explaining the differential distribution of health status by socioeconomic position. This differential distribution is an enduring tendency that holds across a wide variety of health conditions and disease states, from accidents and injuries to mental illness to coronary heart disease, and for a variety of conceptions of socioeconomic position (e.g., social class, minority status, educational attainment) to different degrees at different stages of the life course, and the pattern is generally more pronounced for males than females. The explanatory task of social epidemiology, therefore, is to be able to combine both empirical and theoretical research that links: (1) "cell to society"

(actually, sub-cellular as well) through some sort of "bio-psycho-social translation" or similar alternative (Tarlov 1996); (2) individuals to the experience of a quasi-nested, scaled and stratified social world ("neurons to neighbourhoods to nation-states") (Shonkoff and Phillips 2000); and (3) "cradle to grave" or, stated differently, research that can explain exposures, experiences and outcomes separated by differing, but often very long temporal scales (for example, from early childhood to late adulthood) through different possible modes, including latent, cumulative and pathways effects (Hertzman and Weins 1996). All of this, of course, must account for a complex background of both slowly evolving and punctuated historical changes and an increasingly complex set of geopolitical relations across the globe (Dunn 2006).

## 2.3 Philosophical Underpinnings of Social Epidemiology

Social epidemiology is a science that descends from the positivist tradition. The term positivism has its origins in French social theory. It is often attributed to Auguste Comte (1798–1857), but in fact Comte took the idea from enlightenment thinker Henri Saint-Simon (1760–1825) who was very concerned to establish a new order for society, one based upon the principles of science and industry, as opposed to traditional forms of social domination through nobility and the church.

As a philosophy to guide science, positivism is concerned with the development of knowledge in the form of general statements (ideally universal laws that apply across time and space) obtained through the application of accepted procedures about observable phenomena. Positivism asserts that the only authentic knowledge is that which is based on sense, experience and positive verification.

In its strongest formulation, positivism consists of a set of six principles (Halfpenny 1982; Keat 1981; Bryman 1988; Gartell and Gartell 1996; Ciaffa 1998):

1. The unity of science and its methods – i.e., belief that the methods of the natural sciences are appropriate to the social sciences and indeed, that the only legitimate knowledge of society can come from the application of such methods. Sometimes called methodological naturalism.
2. Knowledge can only be generated by empirical means, specifically, by observation with the senses. This excludes any phenomena that cannot be observed with the senses directly or indirectly with instruments (i.e., metaphysical notions like "social structure" or subjective experience). This principle is sometimes known as empiricism.
3. The objective of scientific research is to explain and predict, and these are symmetrical processes, achieved through the accumulation of empirically established regularities or laws. Most positivists would also say that the ultimate goal is to develop laws of general understanding by creating a model of the phenomenon in question that is simultaneously explanatory and predictive. This principle is often known as inductivism.
4. Scientific knowledge is testable. Research, according to this principle, should be mostly deductive. Scientific theories are seen as providing raw materials for hypotheses, which are developed so that they can be tested in empirical research.

5. Science must distance itself from common sense, which is seen as a source of bias in what is supposed to be an objective activity. Scientific procedures should be capable of generating knowledge that contradicts common sense.
6. Scientific knowledge must be value neutral. The goal of science, in fact, is to produce knowledge regardless of any values, morals or politics held by scientists. The quality of scientific research should be judged by logic, and it should, ideally, produce universal, generalized statements.

There are now numerous distinct versions of positivism, many of which do not subscribe to strong versions of all of these claims. For example, most contemporary social researchers do not believe in the existence of general social laws, and, increasingly, positivist research acknowledges the impossibility of value neutrality. Nevertheless, positivism still constitutes the philosophical basis for social epidemiology. This approach carries with it certain strengths, which are well documented, but also certain weaknesses. In this chapter, the underlying argument is that the realist approach could allow epidemiology's strengths to be maximized and its weaknesses to be minimized.

## 2.4 Key Principles of Critical Realism

Realist philosophy begins from several key propositions, a few of which are worth highlighting from the outset. The first is that the world exists independently of our knowledge of it; that is, there is a real world that exists whether we are aware of it or not (although our access to it is mediated by our senses and other factors) (Sayer 1992). This presupposition is in contradistinction to the viewpoint of idealist philosophy, which says that something only exists if it can be observed. Indeed, the term "realist" is meant to distinguish this school of thought from idealism.

The second key proposition is that all knowledge is fallible and theory laden. Indeed, realists are fond of saying that they are "inherent fallibilists," and that all knowledge is subject to review, change and correction. In this sense, the goal of science for realists is not to discover the truth or universal laws that apply in all times and all places, but rather to develop knowledge that is "practically adequate" but that is constantly subject to revision based on the fallibilist notion above. Practical adequacy will be discussed in more detail shortly. Indeed, as Sayer (1992) argues

> social science has been singularly unsuccessful in discovering law-like regularities. One of the main achievements of recent realist philosophy has been to show that this is an inevitable consequence of an erroneous view of causation. Realism replaces the regularity model with one in which objects and social relations have causal powers which may or may not produce regularities, and which can be explained independent of them.

Consequently, in realism, "less weight is put on quantitative methods for discovering and assessing regularities and more on methods of establishing the qualitative nature of social objects and relations on which causal mechanisms depend" (Sayer 1992).

As mentioned above, in addition to being fallible, all knowledge is theory laden. According to Sayer (1992), "concepts of truth and falsity fail to provide a coherent view of the relationship between knowledge and its object." But knowledge can be checked empirically, and its effectiveness in informing and explaining successful material practice is "not mere accident" (Sayer 1992). Indeed, this is what is intended by the term "practical adequacy."

A third proposition of realism that is important to the following discussion is that "there is necessity in the world; objects – whether natural or social – necessarily have particular causal powers or ways of acting and particular susceptibilities" (Sayer 1992). An important, but under-acknowledged goal of science is to distinguish between necessity (empirical events that occur because of the operation of internal causal mechanisms of objects) and contingency (empirical events that occur because of contingent factors that are unrelated to the operation of mechanisms). For example, it is because of necessary relations (the causal powers and liabilities of gunpowder, steel and human beings) that bullets fired from a gun can cause grave injuries or fatalities, but whether any particular gun is fired and wounds a particular person is a contingent matter.

A fourth proposition is that the "world is differentiated and stratified, consisting not only of events, but objects, including structures, which have powers and liabilities capable of generating events" (Sayer 1992). Most importantly, structures cannot be observed in the strict sense that positivism would insist (in other words, realism accepts metaphysics), and, like many things in the natural and social world, structures do not generate regular patterns of events.

Finally, realism argues that the social sciences, like epidemiology "must be critical of its object. In order to be able to explain and understand social phenomena we have to evaluate them critically." This concept is described more fully in a subsequent section. In the following section, the term "explanation" is recast from its positivist roots as a realist concept, with implications for rethinking social epidemiology.

## 2.5 "Explanation": A Realist Approach

Realism rejects the positivist assertion that prediction and explanation are symmetrical processes (i.e., that the ability of empirical models to predict events subsequently is a necessary condition for explaining the phenomena). Unlike positivism, from the realist perspective, identification of causal relationships *does not* depend on regularity of patterning of empirical events. The causes of World War I can be identified even though it only happened once and therefore does not constitute an empirical regularity. Indeed, in the social sciences regularities of patterning are rare because the objects of study in social science are typically open systems. Where regularities *are* observed in the social sciences, they "do not approach the universality and precision of those available to physicists and astronomers" (Sayer 1992). The reason for this is that in order for regularity of patterning to occur, the following two conditions must be met: first, "there must be no change or qualitative variation

(e.g. impurities) in the object possessing the causal powers if mechanisms are to operate consistently. This is termed by Bhaskar the 'intrinsic condition for closure'" (Sayer 1992). Human beings have the capacity for learning, self-change and even resistance – to health promotion programs, for instance (see Chap. 12) – and therefore this condition is seldom met in the social sciences. The second condition that must be met for regularity of patterning to occur is that:

> the relationship between the causal mechanism and those of its external conditions which make some difference to its operation and effects must be constant if the outcome is to be regular (the extrinsic condition for closure). (Sayer 1992)

According to Sayer (1992), if *both* the intrinsic and extrinsic conditions are met, a *closed system* exists in which regularities may be produced, but, "most systems we encounter violate these conditions in some way and therefore any regularities they produce are at best approximate and short-lived; these are *open systems*" (Sayer 1992, emphasis in original; see also Bhaskar 1975, 1979a, b, 1989; Cloke et al. 1991; Pawson 1989). Now, the social gradient in health is a pattern that has been observed in hundreds of studies, but there is still a great deal of variance in health status that is not explained by socioeconomic position, strongly suggesting that the objects of study (societies) are open systems. It follows that close adherence to any form of positivism will be limited in terms of generating knowledge that can make a difference.

But not only is regularity of patterning of events rare in the social sciences, "patterns of events, be they regular or irregular, *are not self explanatory, but must be explained by reference to what produces them*" (Sayer 1992, emphasis added). This particular point is fundamental to moving forward in the field of social epidemiology. The observation of persistent inequalities in health status is the foundation of social epidemiology, and, to date, attempts to explain the pattern that do not take this observation as naively given are few in number compared to the number of publications that do little more than demonstrate the pattern empirically (Tarlov 1996; Wilkinson 1994, 1996 are exceptions[2]). Traditional approaches based on positivism (e.g., epidemiology) would suggest that more data is needed, with the expectation that eventually a regularity that better explains the phenomenon, or one that gives more certainty to an empirical relationship, will be found. According to the realist view, however, this relationship is not self-explanatory, but must be explained with reference to what produces it. Such an explanation would be expressed in terms of the mechanisms and structures thought to be responsible and, to use Sayer's (1992) terms, their "tendencies" to act in particular ways, their "ways of acting" or their "causal powers and liabilities." Human beings, for example, are vulnerable to suffocation or contagion of disease (a necessary relation), but who suffocates or contracts a contagious disease is a contingent matter.

[2] Link and Phelan's (1995) work on "fundamental causes" is a step in the right direction of deeper explanations, but I argue that the emphasis on social conditions and socioeconomic position is not fundamental enough, as its explanatory focus is situated primarily at the level of events (as opposed to mechanisms and structures – see Fig. 2.1) and it is not successful in distinguishing between necessary and contingent relations.

## 2.6 Causal Analysis

"To ask for the cause of something," argues Sayer, "is to ask what 'makes it happen', what 'produces', 'generates', 'creates', or 'determines' it, or, more weakly, what 'enables', or 'leads to' it" (1992). These common metaphors for describing causality are insufficient for scientific explanation, however, Sayer argues. According to the realist view, "causality concerns not a relationship between discrete events,… but the *'causal powers'* or *'liabilities'* of objects or relations, or more generally, their ways-of-acting or *'mechanisms'*" (Sayer 1992, emphasis in original). People have the causal powers of being able to work, speak, reason, walk, reproduce, learn, *et cetera*, and have causal liabilities such as susceptibility to persuasion, extremes of temperature, piercing by knives, *et cetera*. Causal powers often

> inhere not simply in single objects or individuals, but in the social relations and structures which they form. Thus the powers of a lecturer are not reducible to her characteristics as an individual but derive from her interdependent relations with students, colleagues, administrators, employer, spouse, etc. (Sayer 1992)

If causation is not a matter of linking discrete events, but rather a matter of causal powers and liabilities, it follows that "powers and liabilities can exist whether or not they are being exercised or suffered; unemployed workers have the power to work even though they are not doing so now" (Sayer 1992). It follows further – and this is a fundamental difference from positivistic methods and reasoning – that "a causal claim is not about 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" (Sayer 1992).

Causal analysis, therefore, is about getting beyond the simple recognition that something produces some change, to understanding what it is about the object(s) that enables it (them) to do this. Often, however, relatively little is known about the mechanisms responsible for processes, for instance in the case of gravity, or the relationship between people's intentions and their actions (Sayer 1992). In the example of the electrical conductivity of copper, however, we do have knowledge of *how* the process works (presence of free ions in its structure) without reference to a "black box" – an actual causal mechanism has been postulated/identified, and it has been shown, at least to date, to be "practically adequate" for most purposes, scientific and otherwise (Sayer 1992). The "mode of inference" by which "events are explained by postulating (and identifying) mechanisms which are capable of producing them is called '*retroduction*'" (Sayer 1992, emphasis in original). Retroduction, it would seem, is crucial to the future of social epidemiologic research. Despite this, the depth with which epidemiological studies published in the top journals engage in retroduction is very shallow.

One final point on causation deserves attention before moving on to discuss types of research and the understanding of "explanation" from the realist perspective. Sayer (1992) reminds us that

> processes of change usually involve several causal mechanisms which may be only contingently related to one another. Not surprisingly then, depending on conditions, the operation

> of the same mechanism can produce quite different results and, alternatively, different mechanisms may produce the same empirical result. At one level this seems unexceptional, although it does not rest easily with the orthodox view of causation in terms of regular associations.

As the analysis of the influences of social and economic factors upon health status at the individual and contextual levels involves a wide diversity of objects and relations stretched across time and space, it is going to be a complicated process that will belie orthodox conceptions of "proof" and certainty. Specifically, it demands identification of causal powers and liabilities of a diverse array of objects and relations and their conjunctures and will require different types of research and ways to integrate them.

But this raises an important question for the limits of epidemiology. As Sayer (1992) puts it, "can generalization and the search for regularities ever assist causal analysis?" He argues that in some cases "the discovery of empirical regularities may draw attention to objects whose causal powers might be responsible for the pattern and to conditions which are necessary for their existence and activation," providing an important role for standard epidemiologic methods. But, he argues, in order to confirm that such patterns point to the operation of causal mechanisms, "qualitative information is needed on the nature of the objects involved and not merely more quantitative data on empirical associations." In epidemiology, more data is routinely recommended in cases where uncertainty still exists, or, alternatively, an author may recommend improved measurement of "exposure" variables. However, more data without a better account of causal mechanisms is unlikely to reduce uncertainty and will do little to enhance explanation. As Sayer (1992) observes:

> So, for example, in epidemiology, ignorance of the causes and conditions of certain diseases may require a resort to mapping and charting quantitative data on a wide range of possible factors. It may seem reasonable to search for a factor which is common to all instances of the disease and hypothesize that this is the cause, or else a factor which is only present where the disease occurs. While they are worth trying, both methods fail to address the problem of finding a *mechanism* which *generates* the disease, as opposed to a factor which merely covaries with it. The weakness of the search for mere associations is illustrated in the well-known story of the drunk who tried to discover the causes of his drunkenness by using such methods: On Monday he had whisky and soda, and Tuesday gin and soda, on Wednesday vodka and soda and on other nights when he stayed sober, nothing; by looking for the common factor in the drinking pattern for the nights when he got drunk, he decides the soda water was the cause. Now the drunk might possibly have chosen alcohol as the common factor and hence as the cause. However, what gives such an inference credibility is not merely the knowledge that alcohol was a common factor, *but that it has a mechanism capable of inducing drunkenness*. (emphasis added)

Although this example may seem trivial, it is of critical importance to the future of social epidemiology. Given that one of the most common modes of analysis in social epidemiology is to identify common and distinguishing properties, it is essential to recognize the limitations of this approach. One obvious limitation is that it depends on the existence of counterfactuals, which might not always exist, and the search for which might miss important knowledge. There are two helpful examples

of this. First is what Oakes (2004) calls "structural confounding," whereby in multilevel analysis there is not a full range of counterfactuals to test hypotheses about the relative influence of the neighbourhood or the individual on individual health (see also Chap. 7). In other words, it is often the case that there are too few poor people living in affluent neighbourhoods and *vice versa* to fully test the relative influence of individual versus neighbourhood income on health. The second example comes from the work of Geoffrey Rose (1985), who points out that typical, individual-based risk factor epidemiology, in its search for commonalities and differences among individuals, cannot account for risks to which whole populations are uniformly exposed.

This is not to say that epidemiology should stop doing such analysis, but as Sayer (1992) says, while many studies

> use available qualitative and causal knowledge to narrow down the list of possible factors to those which might have relevant powers and liabilities… *all too often the qualitative investigation is abandoned just at the point when it is most needed – for deciding the status and the causal (as opposed to statistical) significance of whatever patterns and associations are found*. When this happens, research may occlude rather than reveal causality. (Sayer 1992, emphasis added)

A good example of the kind of causal analysis that arguably needs to be encouraged in social epidemiology is illustrated by recent research by Mark Hunter. Working in South Africa, Hunter (2007) seeks to explain HIV infection rates in informal settlements in the early post-apartheid era via three interlinked dynamics critical to understanding the pandemic: (1) rising unemployment and social inequalities that leave some groups, especially poor women, extremely vulnerable; (2) greatly reduced marital rates and the subsequent increase of one-person households; and (3) rising levels of women's migration, especially through circular movements between rural areas and informal settlements in urban areas. He significantly improves upon previous explanations that lean on a specific culture of sex in South Africa and on the explanation of male migratory patterns and use of prostitutes. Specifically, Hunter identifies connections between poverty, the informal economy, money/sex exchanges and even the changing nature of love relationships in contemporary South African culture to explain HIV rates in these informal settlements where rates are said to be double the national average. To do this, he uses a combination of demographic, ethnographic and historical approaches to get beyond superficial explanations seen in strictly quantitative models focusing on strictly demographic factors. Although not explicitly approached from a realist perspective, Hunter's study illustrates a deeper notion of the term "explanation" than is typical in social epidemiology and reveals causal mechanisms with more detail and complexity, opening up more opportunities for intervention.

One of the vexing things about realist philosophy in science is that it is difficult, on the surface, to distinguish it from conventional social scientific research. It is rare that someone will say that they are doing a "realist analysis" of a phenomenon in the way that they would say they are practicing "realist synthesis" (see Chap. 11) or "realist evaluation" (see Chap. 12). In part this is because realism has penetrated

into the core way of doing things for many social sciences,[3] and also because, unlike in systematic reviews or randomized trials, it is impossible to provide a standardized, formulaic way of doing realist-inspired science. Indeed, one of the challenges of adopting a realist approach is that it can look like it simply involves adopting "mixed methods." But realist science involves much more than that, as it is focused on identifying causal mechanisms and distinguishing between necessary and contingent relations. To learn to "do" realist social science, in other words, it is necessary to learn and apply its principles, as opposed to implementing some formulaic sequence of methods.

## 2.7 Types of Research

If explanation is not reducible to the establishment of statistical associations and predictive models, it follows that we must ask: what is the alternative? For Sayer (1992), "causal analysis is usually closely tied to abstraction and structural analysis and hence explanation to description" of the qualitative aspects of objects, relations, and processes. In some instances,

> the discovery of empirical regularities may draw attention to objects whose causal powers might be responsible for the pattern and to conditions which are necessary for their existence and activation. But in order to confirm these, qualitative information is needed on the nature of objects involved and not merely more quantitative data on empirical associations. (Sayer 1992)

Indeed, it is common in epidemiology to employ strategies to identify distinguishing or common factors in cases of diseases, and to conjecture causes from this information, but this fails to account for the mechanism that *generates* the disease and can lead to false and simplistic conclusions (for difficult and complex questions). One of the important activities of social science, therefore, is to identify those mechanisms, as discussed above, through retroduction. But the tools of retroduction, so to speak, are not obvious. According to the realist view, these tools are abstractions, and these are fundamental to explanation.

In common parlance, the adjective "abstract" often means vague, esoteric or removed from reality. The sense in which the term "abstraction" is used in research is different. According to Sayer (1992),

> an abstract concept, or an abstraction, isolates in thought a *one-sided* or partial aspect of an object. What we abstract *from* are the many other aspects which together constitute *concrete* objects such as people, economies, nations, institutions, activities and so on. (emphasis in original)

[3] Indeed, one of the criticisms of realism when people started writing about it in the social sciences is that many said "we're all realists now" (Cloke et al. 1991). Essentially, it codified the way many social scientists already worked. Even if this critique of unoriginality is valid, it is certainly advantageous that social scientists practice their craft in a way that is self-consciously aware of the philosophical underpinnings, and there is still plenty of relatively superficial analysis that could benefit from the insights of realist philosophy.

Abstract concepts need not be vague: abstractions such as temperature, gender, *et cetera*, are quite concrete and precise. Nor should the notion of abstract be equated with unreality or something residing only in thought – abstractions can and do refer to things that are real (Sayer 1992). "Class," "status," "identity," "neighbourhood," "context," "government" and even "society" are all abstractions that identify fundamental dimensions of human existence and have a certain degree of practical adequacy for explaining them.

On the other side of things, the concept of "concrete objects" "does not merely concern 'whatever exists' but draws attention to the fact that objects are usually constituted by a combination of diverse elements or forces" (Sayer 1992). A person, for instance, combines influences and properties from a wide range of sources (e.g., physique, personality, intelligence, attitudes, etc.), each of which "might be isolated in thought by means of abstraction, as a first step towards conceptualizing their combined effect" (Sayer 1992). As a consequence, understanding concrete events or objects involves a double movement:

> Concrete => abstract, abstract => concrete. At the outset our concepts of concrete objects are likely to be superficial or chaotic. In order to understand their diverse determinations we must first abstract them systematically. When each of the abstracted aspects has been examined, it is possible to combine the abstractions so as to form concepts which grasp the concreteness of their objects (Sayer 1992).

This activity forms the basis for research and, indeed, the relationship between empirical events, mechanisms and structures, on the one hand, and abstract and concrete research on the other, as shown in Fig. 2.1. The figure also shows different types of research and what sorts of activity they are concerned with. Orthodox empirical research operates only at the level of empirical events, seeking to make generalizations (extensive research). Abstract research "deals with the constitution and possible ways of acting of social objects" (Sayer 1992). Concrete research seeks to link abstractions and their concrete referents. Interpretive understanding is "presupposed in all of these types of research" (Sayer 1992), including the final type, synthesis research. Synthesis research seeks to combine all of these elements in a robust way (and is described more fully in Chap. 11).

## 2.8 From Qualitative/Quantitative Research to Extensive/Intensive Research

The previous section appears to advocate for qualitative research to inform quantitative research, or mixed methods. However, the distinction between qualitative and quantitative is of limited utility, as is the term "mixed methods." This distinction reflects an unfortunate tendency for people to acquire training and experience in either qualitative or quantitative methods, and then choose their research problems and questions on the basis of how amenable they are to study with the methods they know. Receptivity to mixed methods is a view that is often seen as progressive. This view is sometimes presented as if the mere mixing of methods *necessarily* produces

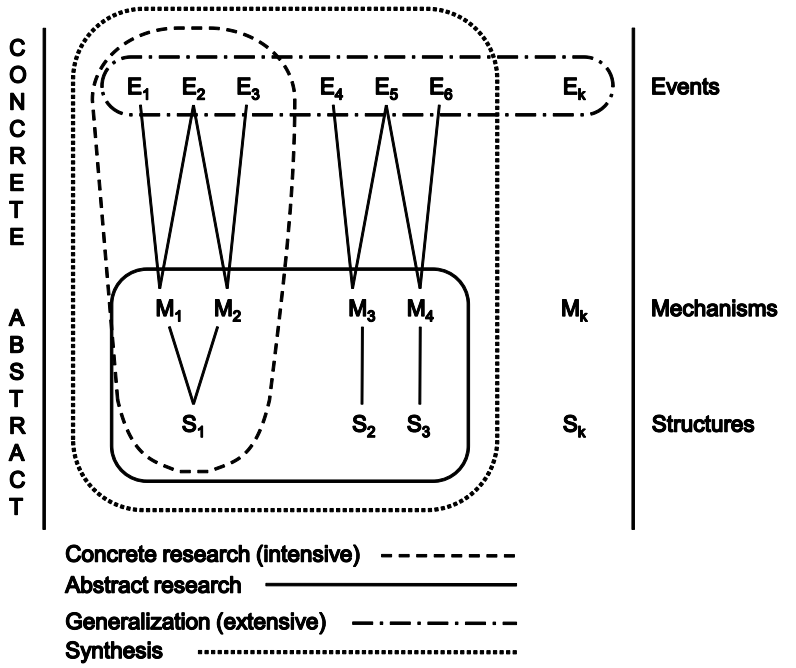


**Fig. 2.1** Types of research (Reprinted from Sayer (1992). With permission from Taylor & Francis Books UK)

better research outcomes. Instead, the appropriateness of the method depends on the type of object and the question one is asking.

According to realism, it is preferable to distinguish between intensive and extensive research rather than qualitative and quantitative research, as intensive and extensive research designs are question driven (rather than method driven) (Sayer 1992; Harré 1979). Intensive research and extensive research differ in a number of their properties (Table 2.1). The chief difference is the kinds of questions that they allow researchers to ask about the phenomena under study. In extensive research, which often involves obtaining relatively superficial information on large numbers of people, questions relating to regularities, common patterns and distinguishing features of a population can be asked. It is also possible to ask how widely certain characteristics or processes are distributed or represented. Often this involves the use of surveys or other kinds of large, quantitative data sets. In extensive research, causal processes cannot be directly observed, as the only relations that can be observed are what Sayer (1992) calls "formal relations of similarity" among "taxonomic groups" (e.g., women under 45 years old). Intensive research, on the other hand, asks questions about how a process works in a small number of cases, often with identifiable people and/or institutions, by identifying substantial relations of connection between such factors as the reasons people give for their actions, their biographies and contingent factors.

**Table 2.1** Intensive and extensive methods

| | Intensive | Extensive |
|---|---|---|
| Research question | • How does the process work in a particular case or a small number of cases?<br>• What did the agents actually do? | • What are the regularities, common patterns, distinguishing features of a population?<br>• How widely are certain characteristics or processes distributed or represented? |
| Relations | • Substantial relations of connection | • Formal relations of similarity |
| Types of groups studied | • Causal groups | • Taxonomic groups |
| Type of account produced | • Causal explanation of the production of certain objects or events, though not necessarily representative ones | • Descriptive "representative" generalizations lacking in explanatory penetration |
| Typical methods | • Study of individual agents in their causal contexts, interactive interviews, ethnography<br>• Qualitative analysis | • Large-scale survey of population or representative sample, formal questionnaires, standardized interviews. Statistical analysis |
| Limitations | • Actual concrete patterns and contingent relations are unlikely to be "representative," "average" or generalizable to other contexts, as they are necessary features of these objects | • Although representative of a whole population, they are unlikely to be generalizable to other populations at different times and places. Problems of ecological fallacy in making inferences about individuals. Limited explanatory power |
| Appropriate tests | • Corroboration | • Replication |

Reprinted from Sayer (1992). With permission from Taylor & Francis Books UK

In intensive research, the groups studied are causal groups that may provide compelling accounts of causal mechanisms (or necessary relations) in real context. The generalizability of such mechanisms to other contexts, however, may be difficult to establish with certainty. This is not necessarily a weakness of intensive methods, as such mechanisms cannot be abstracted from extensive research. When the generalizability of a mechanism is in doubt, then further extensive research can be complementary. On the other side of the ledger, while extensive studies can give representative accounts of phenomena, these tend to lack explanatory penetration. Of course, different kinds of tests are appropriate for each kind of research as well. For intensive research, corroboration is the best way to establish rigour, while in extensive research it is replication (Sayer 1992).

Hunter's (2007) study of the social, economic, cultural and sexual factors that produced specific patterns of HIV transmission in South Africa in the early 2000s is an excellent example of the complementary use of both intensive and extensive methods (although primarily intensive). Hunter also successfully engages in retroduction by abstracting from events the mechanisms and structures that give rise to those events through a rich, complex but cogent account of an important phenomenon – more than would have been possible with extensive methods alone.

## 2.9 Practical Adequacy

One of the questions often raised when methods other than those with roots in positivism are suggested is: how will we know if the results of this research are valid if they are not conducted with objective, scientific methods? Realism offers a powerful and thoughtful perspective on this question through the notion of "practical adequacy," a substitute and, in many ways, a more powerful alternative to notions of truth. "[T]o be practically adequate," argues Sayer (1992), "knowledge must generate expectations about the world and about the results of our actions which are actually realized…" and it must be "…intersubjectively intelligible and acceptable in the case of linguistically expressed knowledge." Quite literally, according to this view, the adequacy of knowledge must be evaluated for its ability to guide practice (i.e., to accomplish some end) in a way that is sensitive to context. Moreover, to "acknowledge that a theory 'works' or has some practical adequacy in a particular context is not to suppose that every one of its constituent elements is 'true' or practically adequate" (Sayer 1992). This notion implies that because of the differentiated nature of the world, we can expect our knowledge of certain processes or phenomena can be unevenly developed yet practically adequate (Sayer 1992). The notion of practical adequacy steps outside conventional notions of truth and falsity, which is helpful partly because it avoids "giving the impression that to hold such false beliefs is necessarily to know nothing and hence to be able to do nothing" (Sayer 1992).

Indeed, the intention and ability to use knowledge to guide practice is an important aspect of the social role of social science. Sayer (1992) suggests that "social science must stand in a critical as well as an explanatory and interpretive relationship to its object and to common-sense knowledge." This implies that

> social science should not be seen as developing a stock of knowledge about an object which is external to us, but should develop a critical self-awareness in people as subjects and indeed assist in their emancipation. It does this first by remembering that its 'object' includes subjects, that the social world is socially produced and hence only one of many possible human constructions. It encourages emancipation and self-development by bringing to light formerly unrecognized constraints on human action. In capitalist societies, with their extraordinarily extended economic relations between anonymous people, the results of people's actions – their own products – take on 'nature-like' qualities in the sense that they react back on us as *blind forces* to which we must submit. (Sayer 1992, emphasis in original)

Taking the world as given, "nature like" or as if "what is, must be" is to naturalize social phenomena and to let common sense guide practice. This is highly problematic because, as Sayer (1992) suggests,

> [a] social science which builds uncritically on common sense, and reproduces these errors, may, at a superficial level, appear to produce correct results. On the other hand, from the standpoint of common sense, which takes its knowledge to be self-evident and beyond challenge, the knowledge produced by critical theories such as marxism will appear to be false because it *conflicts* with what it judges to be the case ("an affront to common sense!"). Yet such theories aim not just to present an alternative or to reduce the illusions inherent in social understanding, but to represent and explain what actually exists as authentically as possible. (emphasis in original)

Indeed, this passage points to the problem with many of the conclusions about health inequalities emerging from social epidemiology, some of which are strongly dependent on unexamined and unacknowledged common sense understandings of the nature of society. The possibilities for a richer theoretical treatment of health inequalities lie not just in the development of an alternative hypothesis, but also in the ability to represent and explain what actually exists in an authentic way that opens up some new possibility for intervention or change in the practices that create the differential distribution of health status. But even more boldly, a further possible outcome of a theoretically richer social epidemiology is a change in society's self-understanding about social inequality and the differential distribution of health status. This is not as outrageous as it may sound if it is recognized that

> *part of "the facts" about human existence is that it depends considerably on societies' self-understanding*, that it is socially produced, albeit only partly in intended ways, and that *changes in this self-understanding are coupled with changes in society's objective form.* (Sayer 1992, emphasis added)

According to this view, "it becomes possible to see how knowledge can simultaneously be not only explanatory and descriptive but also evaluative, critical and emancipatory" (Sayer 1992). That a society's objective form and its self-understanding are interdependent suggests the potential power and impact of theoretical research, if those insights can be successfully disseminated and translated into practice. These are important challenges for the future of social epidemiology. Social change as a result of the permeation of concepts from the social sciences into everyday knowledge is possible, owing to social science's unique relationship to its object of study – an internal one. Concepts from the social sciences differentially impact upon society's self-understanding and do so only in partly intended ways (allowing for the easy dodge of accusations of social engineering – a ridiculous concept on this view), but it is clear that they do have this impact, and often very quickly. Take, for example, Freudian psychology, which emerged in the late 1920s, only to permeate everyday understandings of internal family relations (Oedipus complex) and personality traits (ego) in the western world within a 40–50 year period.

Most individual papers or studies in social epidemiology are relatively limited in their scope with regards to providing a richly theoretically informed explanation of the phenomenon. In part this is because there are aspects of the phenomena that cannot be addressed by the data collected in any one study, so they must be supplemented by secondary reports of other studies and by retroduction – the abstraction of the causal powers and liabilities of the objects under study. To combine such knowledge is a challenging process requiring a careful assessment of the strength of evidence (in the strict empirical sense) and the cogency of theoretical perspectives on the structures and mechanisms that the empirical events are an expression of. And this is the critical point about developing an explanation of this type: the empirical events and data representing them, in addition to being scrutinized for their strength in the traditional positivist sense, must be understood as the concrete manifestation of a number of contingently-related and imperfectly-understood

structures and mechanisms whose causal powers and liabilities have been identified through abstract theoretical research. It is this kind of knowledge, not simply evidence of empirical associations between explanatory and outcome variables that will provide new openings and opportunities for social action to reduce inequalities in health status.

## 2.10 Implications for Social Epidemiology Research

The scope and complexity of the realist perspective makes it difficult to explain in anything less than a book-length project. Nevertheless, there are a few lessons for social epidemiology that can be drawn from the foregoing. These, I would suggest, provide a compelling direction for "rethinking social epidemiology."

The first implication is that the greatest challenge for social epidemiology is not to achieve more accurate measurement of variables, nor is it a need for better or more complete data, or more sophisticated statistical techniques. These kinds of innovations are welcome, but greater precision will not lead to greater certainty, better explanation or greater impact. The most important challenge for social epidemiology arising from the realist perspective is that more attention needs to be paid to the theoretical accounts of causal mechanisms and the distinction between necessary and contingent relations. At some level, this rethinking is going to require more comfort with uncertainty and a retreat from the standards imposed by such positivist cornerstones as replication, hypothesis testing and generalizability. Indeed, if real world impact is the aspiration of social epidemiologists, I would suggest that the search for truth and certainty would be productively replaced by a concern for practical adequacy and the ongoing, reflexive search for fallibility in our knowledge so that it can be refined. By no means does this imply a reduction in rigour, merely a recasting of it.

If we are to provide richer explanations of the phenomena we study, then social epidemiology needs to recognize that causality can never be decided by empirical research alone. It must be expressed in terms of causal powers and liabilities of objects, and in the description of necessary and contingent relations. In order for more social epidemiologists to embrace the kind of research illustrated by Hunter's (2007) work, some institutional change is required. The careers of social epidemiologists will have to be measured less on the number of publications they produce and the impact factor of the journals in which they appear, and more on other factors, such as greater credit for writing books, that would allow social epidemiologists to engage in more exploratory research that would provide richer explanations of social epidemiologic phenomena. Moreover, we need publication outlets that accept longer articles that permit more in-depth exploration of the theoretical and conceptual bases of our work, as opposed to the few paragraphs of speculative interpretation that we usually get at the end of articles in typical epidemiological journals.

In order to aid investigations of richer accounts of causal mechanisms, a more nuanced understanding of the value of various kinds of methods is required. I strongly suggest that we eschew the distinction between qualitative and quantitative methods in favour of the distinction between intensive and extensive research. If more social epidemiologists and social scientists of health had a greater understanding of this distinction, then specialists in both qualitative and quantitative methods would have a better understanding of the relative strengths and weaknesses of their own approaches and also their complementarity.

This raises a critically important implication for training: learning how to execute the methods, the cornerstone of most graduate-level epidemiology programs, is simply inadequate. It is imperative that students understand where methods come from, both historically (i.e., positivism as part of the enlightenment era and as a reaction to the hegemony of the church) and in terms of their philosophical underpinnings. Use of a realist perspective can be helpful in this regard, given how research decisions that need to be made must be made on an appeal to the key philosophical underpinnings.

To embrace a realist perspective and deeper inquiry into causal mechanisms and structures will require greater openness to experimentation with methodology, including natural experiments, ethnographic methods, *et cetera*, because these methods tell us how causal mechanisms work in context. Experimentation is especially valuable because a system behaves differently when it is perturbed than when it is stable, and this perturbation can reveal dormant causal mechanisms in a system. Observational research, therefore, has critical limitations, so natural experiments are needed and welcomed. The non-random assignment of subjects and other messy aspects of such research, rather than things to be disdained, are useful, as they provide learning opportunities. That said, social epidemiology has to embrace designs and methods that can capture much of the qualitative information about the messiness of the implementation of interventions and not the typically clean and sterile "did this intervention have an effect" research (see Chap. 12).

Social epidemiology is growing quickly and has never been more widely read by academics, practitioners and the general public. The realist perspective has a great deal to offer in transforming this popularity into impact. This theme is present in a number of the chapters that follow in this book, and hopefully it can be embraced.

## References

Bambra C, Pope D, Swami V et al (2009) Gender, health inequalities and welfare state regimes: a cross-national study of thirteen European countries. J Epidemiol Community Health 63:38–44

Bhaskar R (1975) A realist theory of science. Leeds Books, Leeds

Bhaskar R (1979a) The possibility of naturalism. Harvester, Brighton

Bhaskar R (1979b) On the possibility of social scientific knowledge and the limits of naturalism. In: Mepham J, Ruben D-H (eds) Issues in Marxist philosophy, vol 3, Epistemology, science, ideology. Humanities Press, Atlantic Highlands

Bhaskar R (1989) Reclaiming reality. Verso, London
Bryman A (1988) Quantity and quality in social research. Routledge, London
Ciaffa JA (1998) Max Weber and the problems of value-free social science: a critical examination of the Werturteilsstreit. Bucknell University Press, Lewisburg
Cloke P, Philo C, Sadler D (1991) Approaching human geography: an introduction to contemporary theoretical debates. Guildford Press, New York
Dunn JR (2006) Speaking theoretically about population health: commentary on carpiano RM and daley DM "A guide and glossary on theory-building for population health". J Epidemiology & Community Health 60:572–573
Forbes A, Wainwright SP (2001) On the methodological, theoretical and philosophical context of health inequalities research: a critique. Soc Sci Med 53:801–816
Gartell CD, Gartell JW (1996) Positivism in sociological practice: 1967–1990. Can Rev Sociol 33:143–158
Halfpenny P (1982) Positivism and sociology: explaining social science. Allen and Unwin, London
Harré R (1979) Social being: a theory for social psychology. Blackwell, Oxford
Hertzman C, Marmot M (1996) The leading hypothesis and its discontents: a synthesis of evidence and outstanding issues regarding the east-west life expectancy gap. In: Hertzman C, Kelly S, Bobak M (eds) East-West life expectancy gap in Europe. Kluwer, Dordrecht
Hertzman C, Weins M (1996) Child development and long-term outcomes: a population health perspective and summary of successful interventions. Soc Sci Med 43:1083–1095
Hunter M (2007) The changing political economy of sex in South Africa: the significance of unemployment and inequalities to the scale of the AIDS pandemic. Soc Sci Med 64:689–700
Kawachi I (1999) Social capital and community effects on population and individual health. Ann NY Acad Sci 896:120–130
Kawachi I, Kennedy BP, Lochner K et al (1997) Social capital, income inequality, and mortality. Am J Public Health 87:1491–1498
Kawachi I, Kim D, Coutts A et al (2004) Commentary: reconciling the three accounts of social capital. Int J Epidemiol 33:682–690
Keat R (1981) The politics of social theory: Habermas, Freud, and the critique of positivism. University of Chicago Press, Chicago
Krieger N (1994) Epidemiology and the web of causation: has anyone seen the spider? Soc Sci Med 39:887–903
Krieger N (1999) Embodying inequality: a review of concepts, measures, and methods for studying health consequences of discrimination. Int J Health Serv 29:295–352
Krieger N (2001) Theories for social epidemiology in the 21st century: an ecosocial perspective. Int J Epidemiol 30:668–677
Link B, Phelan J (1995) Social conditions as fundamental causes of disease. J Health Soc Behav 35:80–94
Muntaner C (2004) Commentary: social capital, social class, and the slow progress of psychosocial epidemiology. Int J Epidemiol 33:674–680
Navarro V, Muntaner C, Borrell C et al (2006) Politics and health outcomes. Lancet 368:1033–1037
Oakes JM (2004) The (mis)estimation of neighborhood effects: causal inference for a practicable social epidemiology. Soc Sci Med 58:1929–1952
Pawson R (1989) A measure for measures: a manifesto for empirical sociology. Routledge, London
Rose G (1985) Sick individuals and sick populations. Int J Epidemiol 14:32–38
Sayer A (1992) Method in social science: a realist approach, 2nd edn. Routledge, London
Sayer A (1993) Postmodernist thought in geography: a realist view. Antipode 25:320–344
Sayer A (1997) Essentialism, social constructionism, and beyond. Sociol Rev 45:453–487
Shonkoff J, Phillips DA (eds) (2000) From neurons to neighbourhoods: the science of early child development. National Research Council and Institute of Medicine, Washington, DC

Tarlov AR (1996) Social determinants of health: the sociobiological translation. In: Blane D, Brunner E, Wilkinson R (eds) Health and social organization: towards a health policy for the 21st century. Routledge, London

Taylor SE, Repetti RL, Seeman T (1997) Health psychology: what is an unhealthy environment and how does it get under the skin? Annu Rev Psychol 48:411–47

Wainwright SP, Forbes A (2000) Philosophical problems with social research on health inequalities. Health Care Anal 8:259–277

Wilkinson RG (1994) The epidemiological transition: from material scarcity to social disadvantage? Daedalus 123:61–78

Wilkinson RG (1996) Unhealthy societies: the afflictions of inequality. Routledge, New York