Theories for social epidemiology in the 21st century: an ecosocial perspective

Nancy Krieger

Keywords Ecology, epidemiology, gender, inequality, political science, psychosocial, race/ethnicity, racism, social class, social determinants of health, social science, socioeconomic, theory

Both thinking and facts are changeable, if only because changes in thinking manifest themselves in changed facts. Conversely, fundamentally new facts can be discovered only through new thinking.’ Ludwick Fleck (1935) Genesis and Development of a Scientific Fact.1, pp.50–51

‘Once we recognize the state of the art is a social product, we are freer to look critically at the agenda of our science, its conceptual framework, and accepted methodologies, and to make conscious research choices.’ Richard Levins and Richard Lewontin (1987) The Dialectical Biologist.2, p.286

Theory

In social epidemiology, to speak of theory is simultaneously to speak of society and biology. It is, I will argue, to speak of embodiment. At issue is how we literally incorporate, biologically, the world around us, a world in which we simultaneously are but one biological species among many—and one whose labour and ideas literally have transformed the face of this earth. To conceptualize and elucidate the myriad social and biological processes resulting in embodiment and its manifestation in populations’ epidemiological profiles, we need theory. This is because theory helps us structure our ideas, so as to explain causal connections between specified phenomena within and across specified domains by using interrelated sets of ideas whose plausibility can be tested by human action and thought.1-3 Grappling with notions of causation, in turn, raises not only complex philosophical issues but also, in the case of social epidemiology, issues of accountability and agency: simply invoking abstract notions of ‘society’ and disembodied ‘genes’ will not suffice. Instead, the central question becomes: who and what is responsible for population patterns of health, disease, and well-being, as manifested in present, past and changing social inequalities in health?

Not surprisingly, theorizing about social inequalities in health runs deep. One reason is that it is fairly obvious that population patterns of good and bad health mirror population distributions of deprivation and privilege. Comments to this effect can be found in the Hippocratic corpus4 and in early texts of ancient Chinese medicine.5 Shared observations of disparities in health, however, do not necessarily translate to common understandings of cause; it is for this reason theory is key. Consider only centuries of debate in the US over the poor health of black Americans. In the 1830s and 1840s, contrary schools of thought ask: is it because blacks are intrinsically inferior to whites?—the majority view, or because they are enslaved?—as argued by Dr James McCune Smith (1811–1865) and Dr James S Rock (1825–1866), two of the country’s first credentialled African American physicians.6 In contemporary parlance, the questions become: do the causes lie in bad genes?, bad behaviours?, or accumulations of bad living and working conditions born of egregious social policies, past and present?7,8 The fundamental tension, then and now, is between theories that seek causes of social inequalities in health in innate versus imposed, or individual versus societal, characteristics.

Yet, despite the key role of theory, explicit or implicit, in shaping what it is we see—or do not see, what we deem knowable—or irrelevant, and what we consider feasible—or insoluble, literature articulating the theoretical frameworks informing research and debates in social epidemiology—and epidemiology more broadly—is sparse.9-12 In this article, I accordingly note the emergence of self-designated social epidemiology in the mid-20th century, review key theories invoked by contemporary social epidemiologists, and highlight the need for advancing theories useful for the 21st century.

‘Social epidemiology’ gains a name ... 

Building on holistic models of health developed between World War I and World War II13 and on the ‘social medicine’ framework forged during the 1940s,14-16 it is in the mid-20th century that ‘social epidemiology’ gains its name-as-such. The term apparently first appears in the title of an article published by Alfred Yankauer in the American Sociological Review in 1950: ‘The relationship of fetal and infant mortality to residential segregation: an inquiry into social epidemiology’,17 a topic as timely now as it was then; Yankauer later becomes editor of the American Journal of Public Health. The term then reappears in the introduction to one of the first books pulling together the behavioural and medical sciences, edited by E Gartly Jaco.
published in 1958, *Patients, Physicians, and Illness: Sourcebook in Behavioral Science and Medicine,* and is included in the title of Jacob's next book, *The Social Epidemiology of Mental Disorders; A Psychiatric Survey of Texas,* published in 1960. By 1969, enough familiarity with the field exists that Leo G. Reeder presents a major address to the American Sociological Association called 'Social epidemiology: an appraisal'. Defining 'social epidemiology' as the 'study of the role of social factors in the etiology of disease', he asserts that 'social epidemiology ... seeks to extend the scope of investigation to include variables and concepts drawn from a theory'—in effect, calling for a marriage of sociological frameworks to epidemiological inquiry.

Soon thereafter, the phrase 'social epidemiology' catches on in the epidemiological literature. Articles appear with such titles as: 'Contributions of social epidemiology to the study of medical care systems', published by S. Leonard Syme in 1971, and 'Social epidemiology and the prevention of cancer', published by Saxon Graham et al. in 1972. By the end of the century, the first textbook is published with the title *Social Epidemiology,* co-edited by Lisa Berkman and Ichiro Kawachi. Despite these gains, it is nevertheless sobering to realize that among the slightly over 432,000 articles indexed in Medline by the keyword 'epidemiology' between 1966 and 2000, only 4% also employ the keyword 'social', and—as Reeder surely would be sorry to learn—fewer than 0.1% are additionally indexed by the term 'theory'. Clearly, there is room for improvement—and reflection.

**Current theoretical trends in social epidemiology**

Contemporary social epidemiology, however, is not without its theories. The three main theories explicitly invoked by practising social epidemiologists are: (1) psychosocial, (2) social production of disease and/or political economy of health, and (3) ecosocial theory and related multi-level frameworks. All seek to elucidate principles capable of explaining social and biological conditions in shaping population health, how they integrate social and biological explanations, and thus their recommendations for action.

**Psychosocial theory**

First, psychosocial theory. As is typically the case with scientific theories, its genesis can be traced to problems prior paradigms could not explain, in this case, why it is that not all people exposed to germs become infected and not all infected people develop disease. One response, first articulated in the 1920s and refined in the 1950s as epidemiologists increasingly study cancer and cardiovascular disease, is to explain the etiological framework from simply 'agent' to 'host-agent-environment'. Despite conceptual expansion, several restrictive assumptions still pervade the new framework's very language. 'Agency', for example, remains located in the 'agent'—typically an exogenous entity that acts upon a designated 'host'; terminology alone renders it inhospitable to conceive of the 'host' having 'agency'! 'Environment', moreover, serves as a catch-all category, with no distinctions offered between the natural world, of which we humans are a part and can transform, and social institutions and practices which we, as humans, create and for which we can hold each other accountable. Gaining complexity without an explicit accounting of social agency, the model becomes increasingly diffuse and, by 1960, the spiderless 'web of causation' is born.

The importance of the 'host-agent-environment' model for psychosocial epidemiology is evidenced by the title of one of the field's still-defining papers: John Cassel's (1921–1976) final opus, 'The contribution of the social environment to host resistance'. Published in the *American Journal of Epidemiology* in 1976, the year of Cassel's death, this article expands upon frameworks elaborated in the 1940s and 1950s linking vulnerability to disease to both physical and psychological stress. Positing that in 'modern' societies exposure to pathogenic agents is ubiquitous, Cassel argues that to explain disease distribution we must therefore investigate factors affecting susceptibility:

> The question facing epidemiological inquiry then is, are there categories or classes of environmental factors that are capable of changing human resistance in important ways and making subsets of people more or less susceptible to these ubiquitous agents in our environment?

To Cassel, in prosperous nations, relevant modifying factors are unlikely to include 'nutritional status, fatigue, overwork, or the like'. More promising candidates lie in what he calls the 'social environment', comprised of psychosocial factors generated by human interaction. Cassel's central hypothesis is that the 'social environment' alters host susceptibility by affecting neuroendocrine function. His list of relevant psychosocial factors includes: dominance hierarchies, marginal status in society, including social isolation, bereavement, and, as acting as a buffer to all of the above, the 'psychosocial asset' of 'social support'. In Cassel's view, these psychosocial factors, considered together, explain the puzzle of why particular social groups are disproportionately burdened by otherwise markedly distinct diseases, e.g. tuberculosis, schizophrenia, and suicide. Shifting attention from 'specific aetiology' to 'generalized susceptibility'—while acknowledging that what disease a person gets is dependent on prior exposures—Cassel ultimately concludes that, in his own words, the most 'feasible' and promising interventions to reduce disease will be 'to improve and strengthen the social supports rather than reduce the exposure to stressors'. Following Cassel's article, research in psychosocial epidemiology blossoms. Between 1966 and 1974, the keywords 'psychosocial' and 'epidemiology' together index only 40 articles in Medline; between 1995 and 1999 alone, the number jumps to nearly 1200. Indicating new ideas are 'in the air', new polysyllabic terms emerge—such as psychoneuroimmunology, whose proliferating prefixes hint some important concepts have yet to be tabbed down. Fortunately, newer additions gaining currency are appreciably shorter. One is 'allostasis', introduced as an alternative to 'homeostasis' in 1988 by Peter Sterling and Joseph Eyer to describe systems that achieve balance through...
change. Its successor, ‘allostatic load’, is then introduced by Bruce McEwen to describe ‘wear-and-tear from chronic over-activity or underactivity’ of systems ‘that protect the body by responding to internal and external stress’, including ‘the autonomic nervous system, the hypothalamic-pituitary-adrenal (HPA) axis, and cardiovascular, metabolic, and immune systems’. One new implication is that psychosocial stressors can be directly pathogenic, rather than alter only susceptibility. And, consonant with the emerging lifecourse perspective—which holds that health status at any given age reflects not only contemporary conditions but prior living circumstances, in utero onwards—‘allostatic load’ draws attention to long-term effects of both chronic and acute stressors. Other new work extends Cassel’s insights to focus on ‘social capital’ and ‘social cohesion’, which—although defined differently by diverse schools—are construed (and contested) as population-level psychosocial assets which shape population health by influencing norms and strengthening bonds of ‘civil society’.

In summary, then, a psychosocial framework directs attention to endogenous biological responses to human interactions. Its focus is on responses to ‘stress’ and on stressed people in need of psychosocial resources. Comparatively less attention, theoretically and empirically, is accorded to: (1) who and what generates psychosocial insults and buffers, and (2) how their distribution—along with that of ubiquitous or non-ubiquitous pathogenic physical, chemical, or biological agents—is shaped by social, political and economic policies. Time also takes a back seat, in that except for reference to periods of rapid social change, the question of whether changing levels of stress are sufficient to explain secular trends in disease and death receives little attention. It is as if, paraphrasing Aaron Antonovsky’s (1923–1994) penultimate lament, the study of why some people swim well and others drown when tossed into a river displaces study of who is tossing whom into the current—and what else might be in the water. To ask the latter questions, however, brings us to other schools of thought.

**Social production of disease/political economy of health**

A second theoretical framework accordingly introduces agency to the ‘upstream-downstream’ metaphors increasingly invoked in social epidemiology today. Hearkening back to social analyses of health of the 1830s and 1840s, as well as 1930s and 1940s, this school of thought—emerging in the politically turbulent 1960s and 1970s—focuses on what it terms the ‘social production of disease’ and/or ‘political economy of health’.

Articles appear with such titles as: ‘A case for refocusing upstream: the political economy of illness’, ‘The social production of disease and illness’, and—recalling the trend’s Marxist origins and its advocacy of ‘materialist’ analyses of health, even: ‘Hypertension in American society: an introduction to historical materialist epidemiology’. These and kindred papers are published, however, in journals unlikely to be on the regular browsing list of most epidemiologists—for example, the *International Journal of Health Services*, founded in 1971 by Vicente Navarro and *Review of Radical Political Economics*. By 1979, the trend’s broad theoretical contours are encapsulated in two books: *The Political Economy of Health*, by Lesley Doyal, a British health policy analyst, and *Epidemiologia Economia Medicina y Política*, by Jamie Breilh, an Ecuadorian epidemiologist.

Arising in part as critique of proliferating blame-the-victim ‘lifestyle’ theories, which emphasize individuals’ responsibility to ‘choose’ so-called ‘healthy lifestyles’ and to cope better with ‘stress’, these new analyses explicitly address economic and political determinants of health and disease, including structural barriers to people living healthy lives. At issue are priorities of capital accumulation and their enforcement by the state, so that the few can stay rich (or become richer) while the many are poor—whether referring to nations or to classes within a specified country. Recast in this manner, determinants of health are analysed in relation to who benefits from specific policies and practices, at whose cost. Core questions include: how does prioritizing capital accumulation over human need affect health, as evinced through injurious workplace organization and exposure to occupational hazards, inadequate pay scales, profligate pollution, and rampant commodification of virtually every human activity, need, and desire? What, too, is the public health impact of state policies enforcing these priorities?—whether by regulation or de-regulation of corporations, the real estate industry, and interest rates; or by enactment or repeal (or enforcement or neglect) of tax codes, trade agreements, labour laws, and environmental laws; or by absolute and relative levels of spending on social programmes versus prisons and the military; or by diplomatic relations with, economic domination of, and even invasion of countries abroad? The underlying hypothesis is that economic and political institutions and decisions that create, enforce, and perpetuate economic and social privilege and inequality are root—or ‘fundamental’—causes of social inequalities in health. Revisiting issues of agency and accountability, theoretical analyses examine interdependence of institutional and interpersonal manifestations of unjust power relations; resources to counter these adverse conditions are reframed, no longer ‘buffers’ but rather strategies for community (not just individual) ‘empowerment’ and social change.

Within this trend, initial conceptual and empirical analyses chiefly focus on class inequalities in health within and between countries. Related contemporary questions include: what are the health impacts of rising income inequality, of structural adjustment programmes imposed by the International Monetary Fund and the World Bank, of neoliberal economic policies favouring dismantling of the welfare state, or of free-trade agreements imposed by the World Trade Organization? Other analyses address social inequalities involving race/ethnicity, gender and sexuality, as they play out within and across socioeconomic position, within and across diverse societies. Relevant questions include: what are the health consequences of experiencing economic and non-economic forms of racial discrimination—or of men dominating and abusing women—or of civilians and soldiers verbally or physically queer-bashing lesbian, gay, and transgendered people? Recently emerging environmental justice movements likewise bring critical attention to corporate decisions and government complicity in transferring toxic waste to poor countries and to poor regions within wealthy countries, especially poor communities of colour. The call for action premised on this framework is thus, minimally, for ‘healthy public policies’, especially redistributive policies to reduce poverty and income inequality, if not for ‘wider campaigns for sustainable development, political freedom, and economic and social justice’.
Four implications for action accordingly flow from a social production of disease/political economy of health perspective. One is that strategies for improving population health require a vision of social justice, backed up by active organizing to change unjust social and economic policies and norms. Another is that absent concerns about social equity, economic growth and public health interventions may end up aggravating, not ameliorating, social inequalities in health if the economic growth exacerbates economic inequality and if the public health interventions are more accessible and acceptable to affluent individuals. A third is that greater familiarity with the emerging field of health and human rights—supplemented by analyses of who gains from neglecting or violating these rights—is likely to improve the real-world efficacy of social epidemiologists’ work, by providing a systematic framework for delineating governmental accountability to promote and protect health, premised, in the first instance, upon the 1948 Universal Declaration of Human Rights and its recognition of the indivisibility and interdependence of civil, political, economic, social and cultural rights. And fourth, social epidemiologists must be key actors in ensuring viability of the vital public health activity of monitoring social inequalities in health, for without such work—which is our particular job to do—it is impossible to gauge progress and setbacks in reducing social inequalities in health.

Yet, despite its invaluable contributions to identifying social determinants of population health, a social production of disease/political economy of health perspective affords few principles for investigating what these determinants are determining. Biology is opaque. Focusing on relative risks across specified social groups, these analyses rely chiefly on critical appraisals of population distributions of known risk and protective factors, most of which ironically are individual-level characteristics identified by conventional epidemiological research. In the case of breast cancer, for example, analyses might focus on social determinants of a variety of reproductive risk factors (e.g. age at menarche, use of oral contraceptives, age at and number of pregnancies), but would be as constrained as conventional analyses in explaining the portion of cases not attributable to these factors. Nor does an emphasis on ‘fundamental social causes’ offer principles for thinking through, systematically, whether—and if so, which—specific public health and policy interventions are needed to curtail social inequalities in health, above and beyond securing adequate living standards and reducing economic inequality. In the background is Thomas McKeown’s (1914–1988) famous argument that 19th century declines in infectious disease mortality in the UK and US are due chiefly to improved nutrition, not medical interventions. Yet, as Simon Szreter and other public health historians have convincingly demonstrated, McKeown is only half right: although medical care per se can claim little credit for declines in incidence or mortality before World War II, economic growth alone did not improve health. Rather, specific public health policies, e.g. those aimed at cleaning the water and eliminating bovine tuberculosis, were also of fundamental importance.

Stated another way, both improved living standards and non-economic interventions (albeit with economic costs and consequences) matter. Moving from an ‘either/or’ to a ‘both/and’ logic requires multi-level frameworks integrating social and biological reasoning and history, and it is to such new theoretical efforts in social epidemiology—building on prior ideas infused into ‘social medicine’ in the 1940s— that I now turn.

**Ecosocial theory and related multi-level dynamic perspectives**

Perhaps one sign of the ferment in contemporary social epidemiological thought is the fact that pictorial depictions of newer frameworks to explain current and shifting patterns of disease distribution refuse to stay in a single plane. Instead, unlike prior images—whether of a triangle connecting ‘host’, ‘agent’ and ‘environment’, or a ‘chain of causes’ arrayed along a scale of biological organization, from ‘society’ to ‘molecular and submolecular particles’, a spiderless two-dimensional ‘web of causation’, or a ‘causal pie’—the new mental pictures are both multidimensional and dynamic. The terminology, too, is changed, as each invokes literal—and not just metaphorical— notions of ecology, situating humans as one notable species among many co-habiting, evolving on, and altering our dynamic planet. I refer especially to three explicitly named frameworks:

1. **‘ecosocial’ theory**, a term I introduced in 1994, with its visual fractal metaphor of an evolving bush of life intertwined at every scale, micro to macro, with the scaffolding of society that different core social groups daily reinforce or seek to alter;

2. **‘eco-epidemiology’** proposed by Mervyn Susser in 1996, with its image of ‘Chinese boxes’, referring to nested ‘interactive systems’, each with its localized structures and relationships, and

3. the **‘social-ecological systems perspective’** invoked by Anthony McMichael in 1999, depicting a cube, representing the ‘present/past’, whose three axes extend from individual-to-population, proximate-to-distal, static/modular to life course and which is projected forward, to ‘future’.

Their goal is not a totalizing theory to explain everything (and therefore nothing), but rather to generate a set of integral (and testable) principles useful for guiding specific inquiry and action, much as evolutionary theory (broadly writ, with contending interpretations) guides biological disciplines ranging from paleontology to molecular biology. And, specifically in the case of ecosocial theory, its fractal image deliberately fosters analysis of current and changing population patterns of health, disease and well-being in relation to each level of biological, ecological and social organization (e.g. cell, organ, organism/individual, family, community, population, society, ecosystem) as manifested at each and every scale, whether relatively small and fast (e.g. enzyme catalysis) or relatively large and slow (e.g. infection and renewal of the pool of susceptibles for a specified infectious disease).

That each of these frameworks explicitly incorporates the prefix ‘eco’ or term ‘ecological’ in its name is revealing. Ecology, after all, is a science devoted to study of evolving interactions between living organisms and inanimate matter and energy...
over time and space. Core to an ecological approach are concerns with:

1. **scale**: referring to quantifiable dimensions of observed spatio-temporal phenomenon, whether measured in nanoseconds or millennia, microns or kilometres;

2. **level of organization**: theorized and inferred, in relation to specified nested hierarchies, from individual to population to ecosystem;

3. **dynamic states**: reflecting combined interplay of specified animate and inanimate ‘inputs’ and ‘outputs’, with recognition that operative processes and phenomena may be scale-dependent (e.g. factors relevant to self-regulation of an organism’s body temperature differ from those involved in self-regulation of the earth’s global temperature);

4. **mathematical modelling**: employed to illuminate how groupings of organisms and processes work together, using both idealized minimal and detailed synthetic models—both to render complexity intelligible and because large-scale experiments are rarely feasible;

5. **understanding unique phenomena in relation to general processes**: in the case of populations, for example, no two forests are ever identical, yet share important features and processes in common relevant to understanding their genesis, longevity, and degradation or decline.

Recognizing, however, the importance of social, political, and economic processes in shaping epidemiological profiles, two of the frameworks—‘ecosocial’ and ‘social-ecological systems perspective’—additionally explicitly indicate in their very names that ecological analysis is not intended to be a substitute or metaphor for social analysis. Rather, they distinguish ecological theory from the diverse social theories upon which they and the other social epidemiological frameworks rely. In doing so, these frameworks part company with other theoretical perspectives that invoke ‘ecology’ as a metaphor, e.g. ‘social ecology’ and ‘human ecology’, and which employ organic analogies that obscure accountability for social divisions and processes by reinterpreting them as ‘natural’ phenomena (e.g. migration of populations to cities and gentrification recast as analogous to plant succession).

Nascent, these emerging ecologically inclined multi-level social epidemiological frameworks remain rather sketchy, the bare beginnings of a mental map. Much more elaboration is required: calling the question can perhaps spur the needed work. Concomitantly, explicit applications to aetiological inquiry and to interventions are only just underway. From an ecosocial perspective, however, it is possible to formulate several constructs that can begin to serve as a mental checklist for epidemiological research. Focused on the guiding question of ‘who and what drives current and changing patterns of social inequalities in health’, the ecosocial approach (but not necessarily the other multi-level frameworks) fully embraces a social production of disease perspective while aiming to bring in a comparably rich biological and ecological analysis. Relevant ecosocial constructs thus minimally include:

1. **embodiment**, a concept referring to how we literally incorporate, biologically, the material and social world in which we live, from conception to death; a corollary is that no aspect of our biology can be understood absent knowledge of history and individual and societal ways of living;

2. **pathways of embodiment**, structured simultaneously by: (a) societal arrangements of power and property and contingent patterns of production, consumption, and reproduction, and (b) constraints and possibilities of our biology, as shaped by our species’ evolutionary history, our ecological context, and individual histories, that is, trajectories of biological and social development;

3. **cumulative interplay between exposure, susceptibility and resistance**, expressed in pathways of embodiment, with each factor and its distribution conceptualized at multiple levels (individual, neighbourhood, regional or political jurisdiction, national, inter- or supra-national) and in multiple domains (e.g. home, work, school, other public settings), in relation to relevant ecological niches, and manifested in processes at multiple scales of time and space;

4. **accountability and agency**, expressed in pathways of and knowledge about embodiment, in relation to institutions (government, business and public sector), households and individuals, and also to accountability and agency of epidemiologists and other scientists for theories used and ignored to explain social inequalities in health; a corollary is that, given likely complementary causal explanations at different scales and levels, epidemiological studies should explicitly name and consider the benefits and limitations of their particular scale and level of analysis.

With these constructs at hand, we can begin to elucidate population patterns of health, disease and well-being as biological expressions of social relations, and can likewise begin to see how social relations influence our most basic understandings of biology and our social constructions of disease—thereby potentially generating new knowledge and new grounds for action.

Consider, as one example, the phenomenon of pregnancy in relation to risk of cancer. Let us start with breast cancer. As is well known, pregnancy decreases risk of breast cancer over the lifetime if it occurs early, but thereafter increases risk, especially after age 35. This phenomenon is often invoked to explain, in part, why incidence of breast cancer increases with affluence and why the rate has climbed during the 20th century (over and above increases due to earlier age at menarche), since more educated women tend to have children later in life and educational level of women, especially in industrialized societies, has generally been on the rise. Notably, all three social epidemiological frameworks—psychosocial, social production of disease, and ecosocial—would highlight how social conditions, including women’s social status, available birth control technology and access to abortion, affect age at first pregnancy. An ecosocial approach, however, would raise questions beyond social determinants of age at first pregnancy to inquire how pregnancy itself is conceptualized in relation to risk of breast cancer. Constructs of ‘embodiment’, ‘pathways of embodiment’, and the ‘dynamic and cumulative interplay
between exposure, susceptibility and resistance’ would require analysing pregnancy in relation to developmental biology of the breast (especially maturation of lobules and ducts and also altered rates of apoptosis) as well as its effects on the endocrine system (synthesis of hormones within the breast plus alteration in magnitude and frequency of hormonal fluctuations) and cardiovascular system (increased vascularization of the breast).84,104 A concern with ‘accountability and agency’, as well as scale and level, would additionally challenge gender-biased views positing reproductive hormones as primary determinants of women’s health.68–70,105 The net result would be to reconceptualize pregnancy not simply as an ‘exposure’ but also as a biological process capable of altering susceptibility to exogenous carcinogens.84,104,106 This is the thinking, in part, behind new aetiologic research on environmental pollution and breast cancer; although ‘answers’ are not yet in as to causal relationships, at least the question is posed.106–108

Similar fresh and integrative thinking motivates a recent novel study including men and women which asks if relationships between parity and cancer incidence are due to the biology of pregnancy or to other social factors ‘that are influenced by or are influencing family size’.109, p.477 Tellingly, parity is equally associated among women and men with risk of three types of cancer: oral and pharyngeal (reflecting greater use of tobacco and alcohol by childless men and women, a topic itself meriting investigation) and malignant melanoma—for which the parity/risk association had been previously interpreted only in hormonal terms, and only for women. For two sites, however, thyroid and Hodgkin’s disease, parity is associated with incidence only among women. One implication of these findings is not to presume parity exerts effects solely by pregnancy-related biological processes; the other is to consider the social meaning of parity even when the biology of pregnancy is relevant. Simplistic divisions of the social and biological will not suffice.

Consider, too, how an ecosocial perspective can contribute to unravelling the unexplained excess risk of hypertension among African Americans.7 Moving beyond eclectic, purely psychological, or purely economic sets of risk factors, the four ecosocial constructs can systematically be used to propose six discrete—yet entangled—multi-level pathways linking expressions of racial discrimination and their biological embodiment across the life course.7,8,64–65,110 These are:

1. **Economic and social deprivation:** For example, residential and occupational segregation lead to greater economic deprivation among African Americans and increased likelihood of living in neighbourhoods without good supermarkets; risk of hypertension is increased by cheap, high fat, high salt and low vegetable diets; also, economic deprivation increases risk of being born preterm, thereby harming development of kidneys and increasing likelihood of chronic salt retention;7,8,110–112

2. **Toxic substances and hazardous conditions:** Residential segregation increases risk of exposure to lead paint in older houses and to soil contaminated by lead from car exhaust (due to closer proximity of residences to streets or freeways); lead damages renal physiology, increasing risk of hypertension;7,8,110,113,114

3. **Socially inflicted trauma:** perceiving, recalling or anticipating interpersonal racial discrimination provokes fear and anger, triggering the ‘flight-or-fight’ response; chronic triggering of this pathway increases allostatic load, leading to sustained hypertension;7,8,34,35,110,111,115,116

4. **Targeted marketing of commodities:** targeted marketing of high-alcohol content beverages to black communities increases likelihood of harmful use of alcohol to reduce feelings of distress; excess alcohol consumption elevates risk of hypertension;112,117

5. **Inadequate health care:** poorer detection and clinical management of hypertension among African Americans increases risk of untreated and uncontrolled hypertension;112,118

6. **Resistance to racial oppression:** individual and community resources and social movements to counter racism and to enhance dignity, along with enactment and implementation of legislation to outlaw racial discrimination, reduces risk of hypertension among African Americans.7,8,110,119

Embracing social determinants ignored by biomedical approaches, the ecosocial approach thus recasts alleged ‘racial’ differences in biology (e.g. kidney function, blood pressure) as mutable and embodied biological expressions of racism.101 Emphasizing accountability, it extends beyond psychosocial explanations focused on ‘anger’ and ‘hostility’120,121 to the social phenomena—in this case, interpersonal and institutional discrimination—eliciting these responses, as mediated by material pathways. Highlighting dynamic and cumulative interplay between exposure, susceptibility and resistance, it advances beyond social production of disease analyses typically focused on racial/ethnic disparities in socioeconomic position among adults51 to highlight discrimination within class strata plus ongoing biological impact of economic deprivation in early life7,8,36,110 Urging conceptual integration, it advocates coordinated research and action cognizant of the specified multiple pathways and geared to explaining current and changing rates of hypertension, premised on the view that our common humanity demands no less if we are to understand and rectify social inequalities in health.51,122 Thus, more than simply adding ‘biology’ to ‘social’ analyses, or ‘social factors’ to ‘biological’ analyses, the ecosocial framework begins to envision a more systematic integrated approach capable of generating new hypotheses, rather than simply reinterpreting factors identified by one approach (e.g. biological) in terms of another (e.g. social). Suggesting much work remains to be done, however, few of the proposed pathways have been extensively studied and, to date, fewer than 25 epidemiological studies have explicitly investigated somatic consequences of racial discrimination—a mere 0.06% of the nearly 40 000 articles indexed by the keyword ‘race’ in Medline since 1966.7

**Conclusion: theory matters**

In conclusion, theory matters: both to define social epidemiology and to distinguish among trends within this field. These diverse frameworks encourage us to think critically and systematically about intimate and integral connections between our social and biological existence—and, especially in the case of social production of disease and ecosocial theory, to name explicitly who benefits from and is accountable for social inequalities in
health. By focusing attention on under-theorized and under-researched conjunction social and biological determinants of disease distribution, these theories, even in nascent form, can potentially give new grounds for action—and underscore that theory, absent action, is an empty promise.

Ultimately, it remains to be seen whether any of the three theoretical frameworks discussed in this article—psychosocial, social production of disease/political economy of health, and emerging ecosocial and other multi-level frameworks—are best suited for guiding social epidemiological research in the 21st century. If not these theories, however, other frameworks will need to be elaborated to enhance social epidemiologists’ ability to analyse and provide evidence useful for addressing the myriad ways we both embody and transform the co-mingled social and biological world in which we live, love, work, play, fight, ail and die. To generate the data required to test and refine our theoretical frameworks, priority must thus be accorded to: (1) enhanced monitoring of social inequalities of health, so that data are available—cross-stratified—by class, gender, and race/ethnicity and any other social groups subject to economic and social deprivation and discrimination, to gauge progress and setbacks in reducing social inequalities in health, (2) funding interdisciplinary aetiologic research to identify conjoint social and biological determinants of disease at appropriate spatiotemporal scales and levels of organization, and (3) funding interventions based on the findings of this research—with the content of all three priority areas determined by coalitions including sectors of society most burdened by social inequalities in health.

If social epidemiologists are to gain clarity on causes of and barriers to reducing social inequalities in health, adequate theory is a necessity, not a luxury. The old adage still stands: ‘if you don’t ask, you don’t know, and if you don’t know, you can’t act.’ Ultimately, it is theory which inspires our questions, which enables us to envision a far healthier world than the one in which we live, and which gives us the insight, responsibility, and accountability to translate this vision to a reality. Who shall create this theory? The task is ours.

Acknowledgements
Thanks to Sophia Gruskin, Mary Basset, George Davey Smith, and also to two anonymous reviewers, for their helpful comments. No funds from any grant supported this project. A small honorarium, however, was paid by the organizers of the ‘Theory and Action’ conference at which a preliminary version of this paper was first presented.

KEY MESSAGES
- Shared observations of disparities in health do not necessarily translate to common understandings of cause; it is for this reason theory is key.
- In contemporary social epidemiology, the three main theoretical frameworks for explaining disease distribution are: (1) psychosocial, (2) social production of disease/political economy of health, and (3) ecosocial and other emerging multi-level frameworks.
- A psychosocial framework directs attention to endogenous biological responses to human interactions; a social production of disease/political economy of health framework explicitly addresses economic and political determinants of health and disease but leaves biology opaque; ecosocial and other emerging multi-level frameworks seek to integrate social and biological reasoning and a dynamic, historical and ecological perspective to develop new insights into determinants of population distributions of disease and social inequalities in health.
- To gain clarity on causes of and barriers to reducing social inequalities in health, social epidemiologists will need to generate improved theoretical frameworks and the necessary data to test and refine them.

References
6 Krieger N. Shades of difference: theoretical underpinnings of the medical controversy on black-white differences, 1830–1870. Int J Health Services 1987;7:258–79.


NB: Statements pertaining to authorship:

As sole author of this paper, I have: (1) conceptualized and written the entire paper, from start to finish; (2) checked the references for accuracy and completeness; (3) assumed sole responsibility to vouch for its validity; (4) NO conflict of interest, in that I have not received any financial support from any group who will gain from the publication of the paper; (5) NOT published this material previously in a substantively similar form. However, with permission from the IJE editorial office, an earlier version of this paper will appear in a free pamphlet containing the proceeds of a one-day meeting on *Inequality in Health—The Current Debate*, an ESRC-sponsored ALISS Seminar, held jointly with the Academy of Learned Societies for the Social Sciences at the London School of Hygiene on 18 July 2000.