

## Chapter 4

# Fundamental Causality: Challenges of an Animating Concept for Medical Sociology

Jeremy Freese and Karen Lutfey

Arguably, the most important problem at the intersection of sociology and epidemiology is how to understand the pervasive positive relationship between various indicators of social position (hereafter, socioeconomic status or SES) and health. The lower status people are, the sooner they die, and the worse health they have while alive. Negative associations between SES and health overall have been found in almost every place and time for which data permit adequate study, implying that the generalization has held even as the prevalence of particular causes of ill-health and death have varied (see reviews in Marmot 2004; Link and Phelan 1995; Deaton 2002; House et al. 1990). In addition, data suggest that the negative association between at least some indicators of SES and some indicators of health may be increasing in some populations, including the United States (Duncan 1996; Lauderdale 2001; Preston and Elo 1995; Steenland et al. 2004; Krieger et al. 2008). Meara et al. (2008) found that while life expectancy had increased 1.6 years between 1990 and 2000 among those who had attended college, it had not increased at all over this same period among those who had not. While various caveats can be raised, none should detract from appreciating that socioeconomic disparities in health in studied populations overwhelmingly are pervasive and profound.

The obvious scientific question about this inverse relationship between SES and health is “Why?,” but two distinct “Why?” questions exist. First, for any population in which an association between SES and health exists, we can ask why that association exists, there and then. Even if we have complete knowledge of the causal mechanisms responsible for the association within one population, however, that knowledge cannot, by itself, explain why the association extends to other times and places in which the causes of ill-health and mortality differ considerably. Therefore, a second question is why the association persists across populations even as the specific threats to population health change.

Toward addressing the latter, there has been considerable enthusiasm in medical sociology for the proposition that SES is a “fundamental cause” of health. Notwithstanding the contributions of significant precursors (e.g., House et al. 1990; Lieberman 1985), the most prominent and sustained exposition of “fundamental causality” has been by Link and Phelan (1995, 1996, 2000, 2002, 2005); Link et al. 1998; Link 2008; Phelan et al. 2004; Phelan and Link 2005). Although they have articulated the details somewhat differently in different papers, Link and Phelan’s argument consistently emphasizes the intersection of information and resource inequalities for understanding the enduring SES–health relationship.

---

J. Freese (✉)  
Department of Sociology, Northwestern University, 1810 Chicago Ave., Rm. 211,  
Evanston, IL 60208, USA  
e-mail: jfreese@northwestern.edu

To understand their position in brief, consider the following sentence from an article by Sankar et al (2004, p. 2985) in *JAMA*:

Disparities in health status have increased in the United States in the last 50 years despite remarkable advances in our ability to prevent, diagnose, and treat disease.

This sentence places two stylized facts about the recent history of population health in an *ironic* relationship to one another: we know more about how to protect health, and yet inequalities in health outcomes have increased. By contrast, the fundamental cause concept replaces the ironic connection with a causal one – that disparities in health status have increased in the United States in the last 50 years *in significant part because of* remarkable advances in our ability to prevent, diagnose, and treat disease. In other words, work on SES as a fundamental cause of health emphasizes the differential distribution of control over disease and its implications for the resulting distribution of health outcomes. Scientific and technological advances increase health opportunity and those with higher SES benefit more from that opportunity than do others. The fundamental causality literature thereby offers the possibility of a general logic by which the pervasive and enduring character of disparities may be understood as well as the logic for expecting when such disparities would increase or decrease. In doing so, the literature contributes to a greater theoretical understanding of health inequalities than is afforded by studies of specific causes, outcomes, or interventions.

“Fundamental causality” as a concept has informed a wide range of studies in the last decade. As valuable as this work has been, we think that medical sociology and social epidemiology going forward would benefit from increased appreciation of some distinctions and tensions regarding fundamental causality, as these may help animate future inquiry. Accordingly, we attempt to develop a forward-looking articulation of fundamental causality and health disparities from a friendly but critical explication of previously published arguments on this topic. We begin by providing a systematic exposition of Link and Phelan’s arguments about fundamental causality as both conceptual and theoretical contributions; then we consider opportunities for possible synergy among different social science research methods; and finally we consider some implications for policy. Our goal, emphatically, is *constructive*: we seek to provide a theoretical clarification and elaboration which we believe suggests ultimately that the concept of “fundamental causality” may be even more “fundamental” to the sociological agenda for understanding health than has been so far recognized.

### Fundamental Causality as a Concept

Proclaiming that “SES is a *fundamental* cause of health” is not especially interesting if it means only that “SES is a cause of health” or even that “SES is an important and enduring cause of health.” Instead, the adjective “fundamental” must add something specific and meaningful, making “fundamental causality” a particular type of causal relation. One should then be able to articulate the meaning of fundamental causality abstractly, without needing to invoke either “SES” or “health.” Although Link and Phelan have not articulated fundamental causality in such abstract terms, we believe that our formulation in this section is compatible with their reasoning. Briefly articulating “fundamental causality” as a *concept* will also help highlight the distinction between *whether* SES is a fundamental cause of health and any particular explanation of *why* SES is a fundamental cause of health.

First, for X to be a fundamental cause of Y, X has to be a cause of Y. Saying that SES is a fundamental cause of health implies that if individuals’ SES had been different, then their subsequent “life chances” for health outcomes would be different. If the correlation between SES and health was entirely due to poor health causing lower SES, SES would not be a cause of health, much less a fundamental one. Likewise, if the correlation between SES and health was entirely due to some

third variable causing both, such as the unlikely theory that “intelligence” is largely responsible for the apparent causal relationship between SES and health, then SES would not be a cause of health (Gottfredson 2004; Link et al. 2008; Cutler and Lleras-Muney Forth coming).

For X to be a fundamental cause of Y, X must have diverse immediate consequences and Y diverse immediate causes. The many consequences of SES-related resources may influence the many causes of health through a large and complicated series of paths, each of which can be called a *mechanism*. For example, Adler and Newman (2002, p. 66) write, “Low-SES peoples also experience greater residential crowding and noise...Noise exposure has been linked to... hypertension among adults.” If correct, SES differences cause housing differences, which cause noise exposure differences, which cause blood pressure differences, which presumably then cause some increased mortality risk from cardiovascular disease. Even if the ultimate effect is only very slight, this would still be one mechanism linking SES to health. That X has diverse consequences and Y has diverse causes raises the possibility of *massively multiple mechanisms*, a very large number of distinct, specific ways that X and Y are causally connected.

“Fundamental causality” is more compelling as a distinct type of cause if one also stipulates that no single intervening variable accounts for the bulk of the enduring relationship between two variables. For example, if pervasive racial disparities in health were entirely explained by the effects of race on SES and of SES on health, then we would say that race is not itself a fundamental cause of health, but SES (perhaps) would be. Similarly, if the reason SES affected health was dominantly that SES was associated with “stress” and “stress” had various implications for health, then we would see less point to asserting that SES itself was a fundamental cause of health as opposed to just calling attention to the dominant mediating role of stress.

X is not a fundamental cause of Y if there are massively multiple causal mechanisms linking X and Y but they largely cancel each other out. If having higher SES is good for health in many ways, there could still be no association if higher SES was also bad for health in many ways. Instead, then, a fundamental cause relationship implies a *systematic asymmetry* by which the mechanisms overwhelmingly imply an influence of X on Y in one direction rather than the reverse. There may be ways that higher SES is detrimental to health (see, e.g., the discussion of *status pursuit* by Lutfey and Freese 2005, p. 1365), but these must be much weaker in their ultimate consequence than the ways that higher SES promotes health.<sup>1</sup>

The sine qua non of the fundamental cause claim is that this asymmetry in mechanisms is systematically produced, such that, when new mechanisms emerge, they can be expected, more often than not, to preserve the underlying relationship. This distinguishes fundamental causes from other distal causes, because it implies an ultimate limitation to any attempt to “explain” the influence of a fundamental cause solely by reduction to proximate causes. Instead, one must explain also what would warrant the predictive claim that new mechanisms will tend to preserve the relationship between X and Y. We have elsewhere called this a *metamechanism*: an abstract mechanism that explains the generation of multiple concrete mechanisms that reproduce a particular relationship in different places and different times (Lutfey and Freese 2005). The metamechanism provides what we term a *durable narrative* to why the SES–health relationship should be robust to changes in health threats and treatments – an explanation of why a similar association would be observed in diverse socio-historical contexts. In our view, the existence of a durable narrative is what makes fundamental causes “fundamental.”

<sup>1</sup>Some have asserted that gender can also be considered a fundamental cause of health (Graham 2004, p. 112). Surely, it is easy to see many pathways between gender inequalities and health outcomes. However, the greater longevity of the socially disadvantaged group (women) makes us wish for a more detailed explication of the understanding of the fundamental cause concept that yields the assertion that, by simple analogy to SES or race, we can think of gender as a fundamental cause of health.

Regardless of the terminology used, the fundamental cause claim implies not just that mechanisms connecting SES and health typically result in an inverse relationship between the two, but that there are systemic, articulatable reasons why this is so. It is not sufficient, for example, to note the existence of enduring "contextual" features like neighborhood differences, but rather this must be linked to explanation of why wealthier neighborhoods should be, in general, more health-promoting than poorer ones. Systemic explanation is why the fundamental cause concept applied to health can be taken as a challenge to the relentless focus on "risk factors" in epidemiology. *A complete articulation of specific proximate mechanisms of inequality is not a full explanation if it misses an incisive explanation of the mechanisms themselves – incisive in that it makes sense of a diverse set of mechanisms, offers predictive insight into why the population distribution of disease will be surprisingly robust to changes in the causes of ill-health, and calls attention to the possibility of more encompassing interventions.* The notion of fundamental causes allows findings about specific causes and specific disease outcomes to be understood cumulatively in the context of more diffuse, encompassing constructs like socioeconomic status and health.

### Fundamental Causality as a Theory

In articulating their arguments about fundamental causality, Link and Phelan have been engaged primarily with alternatives to the idea that SES is a fundamental cause of health – for example, the assertion that the health gradient is mostly attributable to health causing SES or to stress and other psychophysiological consequences of social hierarchies (Marmot 2004). As useful as this has been, engagement with the issue of whether SES is a fundamental cause of health has resulted in some blurring of theoretical claims about why SES is a fundamental cause of health. Additionally, the primary explanatory concepts are diffuse, and this diffuseness has both virtues and limitations. For future work on health using the notion of fundamental causality to develop most fruitfully, we think distinctions on both these fronts need to be clearer.

### *Differences in Means*

Why do changes in the proximate determinants of health result in new mechanisms that sustain the inverse SES–health relationship in much the same way that the old ones did? Link and Phelan have offered several concise theoretical statements on this question. Consider:

Socioeconomic status operates as a 'fundamental cause' of disease by allowing people with high socioeconomic status to use broadly serviceable resources, such as knowledge, money, and power, to avoid risks and to minimize the consequences of disease once it occurs. (Link and Phelan 1996, p. 599)

SES disparities in mortality arise because people of higher SES use flexible resources to avoid risk and adopt protective strategies. (Phelan and Link 2005, p. 30)

[P]eople with superior resources can use those resources to garner health advantages. (Link and Phelan 2002, p. 732)

[N]ew mechanisms arise because persons higher in SES enjoy a wide range of resources – including money, knowledge, prestige, power, and beneficial social connections – that they can utilize for their advantage. (Link and Phelan 2005, p. 73)

These statements articulate an elegant metamechanism for the pervasiveness of health disparities. Using words like "utilize," "use," "avoid," and "strategies," Link and Phelan direct attention

to the role of individual purposive action, or what they call "health-directed human agency" (Link and Phelan 2002, p. 732). More specifically, they posit SES differences in the *means* of achieving health goals as being the crucial difference by which the fundamental relationship between SES and health is preserved.

If years of health could be bought at auction, presumably the rich would buy more (Goldman and Lakdawalla 2005). If the resources identified with SES confer advantage for actors realizing their preferences for health, then we would expect those with higher SES to have better health. Medical advances have increased the opportunity for health-related agency to yield fruit, thereby allowing differences in access and action to manifest themselves as differences in health outcomes. In principle, the existence of purposive actors with differential means to achieving a broadly valued end is *sufficient* to predict the existence of disparate outcomes, but that implies nothing about the true extent to which purposive action with different means actually leads to observed disparities.

### *The Ambiguity of "Resources"*

As the statements by Link and Phelan above make clear, the workhorse construct for their theorizing about means has been "resources." "Resource" implies agency, a potentiality that can be drawn upon toward furthering ends. In Link and Phelan's formulation, traditional indicators of social standing – namely education, wealth, and occupation – yield heterogeneous resources that purposive actors can use to benefit their health. Material resources like money can be used to secure access to items or services that protect health. Social resources like interpersonal relationships can be used to draw upon to receive access to quality health information or access to providers. Cognitive resources allow individuals to better understand how their actions influence health, to better utilize information sources to protect health, and to better exploit available technologies.

Of these, material resources are most prototypic, but numerous lines of evidence suggest that the importance of specific material resource differences for health disparities may be easily overstated. Increases in population wealth bear an uncertain relationship to population health once state- and institution-level public health changes are taken into account, calling into question how much individuals help their own health by becoming wealthier (Cutler et al. 2006). Various studies by economists have estimated little short- or medium-run gains to individual health from exogenous increases in income (Smith 2007). This creates the possibility for an unfortunate shell game in how researchers think and talk about health disparities: "SES" most immediately evokes income, but education differences are more consequential for health in the United States; "resources" most immediately evokes money, but nonmonetary differences are more important (see Mirowsky and Ross 2003 regarding the "money fallacy"). Deaton (2002, p. 14, 21) goes so far as to call SES "unhelpful" and "useless for thinking about policy" for health disparities because of its vagueness. More pressing for fundamental causality as a theory of health disparity, however, is the question of how far the notion of differences in the agentic use of resources can be stretched and still be useful for explaining the enduring character of disparities.

As elegant as differences in means is as a metamechanism, many of the specific mechanisms invoked by Link and Phelan's arguments and examples do not involve differences in means. They cite health-promoting behaviors whose costs are minimal and for which information about benefits have widely diffused (e.g., "wearing seat belts" [Link and Phelan 2005, p. 74]). They also invoke SES-related circumstances that have implications for health but are not necessarily the result of any personally health-directed effort (e.g., "living in neighborhoods where garbage is picked up often," "having children who bring home useful health information" [Link and Phelan 2000, p. 41, 2002, p. 30, 2005, p. 74]).

The result is a dilemma. On the one hand, purposive action with different means is a clear metamechanism and offers a coherent theoretical narrative of why SES is enduringly related to

health, but also stretches the concept of "resources" to where it fits uneasily at best. We do not regard wearing seat belts as an example of how SES produces differentials in "access to a broad range of circumstances" that promote health (Link and Phelan 2005, p. 74), because there is little reason to think the SES gradient in seat belt usage in the United States has much to do either with "access" to seatbelts (they are legally required in all cars in the United States) or with information about their benefits. On the other hand, when less purposive language is used, arguments may seem to lose the semantic content of a theory altogether. When Phelan and Link (2005, p. 27) state elsewhere that people of higher SES benefit from new health innovations because they can better "harness the benefit" of those innovations, it is unclear what verbs like "harness" mean beyond just saying that people of higher SES benefit more because they benefit more. To be sure, all these examples underscore the distinction between Link and Phelan's position and social selection or stress-centered theories of health disparities. But a theory focused on health-related human agency does not provide a satisfactory explanation of how the fundamental relationship between SES and health is preserved, and this reveals important opportunities for both future theoretical development and empirical research.

### *Complements to Means*

As we have argued, differences in means provide one durable narrative of health disparities, but differences in means among purposive agents do not account for all the cited ways that SES causes health. One way forward is to posit additional metamechanisms of the SES–health relationship that are distinct from differences in means. The relative importance of different metamechanisms in a population is an empirical matter, and one relevant for policy interventions to lower disparities. To this end, we outline three additional metamechanisms here: SES differences in (1) spillovers, (2) habitus, and (3) the ways that social institutions process individuals. In articulating these, we hope also to further clarify the difference between identifying ways that resources may affect health and developing a more comprehensive theory of the pervasive and enduring character of health disparities.

*Spillovers.* Individuals are embedded in social relations in which other people also value their health, and the actions of other people have consequences that accrue differently to people of different social positions. As a result, we might expect that even among high- and low-status individuals who do not especially care about their health, higher SES individuals will have better health because they gain more spillover benefits from the purposive actions of others in their social networks. For example, a business executive who cares less than the average person about her health may still realize health benefits from her choices of job, neighborhood, and social networks, despite none of those choices purposely "utilizing" resources or enacting "strategies" to improve health.

Link and Phelan provide examples of "contextual" effects as support for their position, but they do not articulate a durable narrative for why a decision to live in the most expensive neighborhood one can afford carries health benefits even though the decision itself need not be motivated by any health concerns. Neighborhoods connect individuals to others, many of whom do care, and the differential means by which these others act for their own health – vigilance about local environmental hazards, for instance, or caring about the quality of nearby health services – can have positive spillovers for others to whom they are connected. More generally, such spillovers most affect those to whom one is socially close (e.g., neighbors, family, friends), and social distance is lower for individuals of similar SES. As a result, new knowledge about health confers disproportionate benefit to high-SES individuals independent of the exercise of their own agency.

We suggest that spillovers provide an important route for connecting findings about social network effects to the idea of health as a fundamental cause. Christakis and Fowler (2007, 2008) provide evidence that both becoming obese and smoking may be influenced by having a friend who has

done the same. If so, then becoming obese or smoking have differing social costs depending on social ties as a result of earlier behavior by others. If SES is correlated with social ties, then spillovers provide a metamechanism by which network diffusion can preserve a fundamental relationship between SES and health.

In the same way, spillovers may also help understand the relationship between fundamental causality and proffered mechanisms like lower SES individuals being exposed to "more advertising for tobacco and alcohol" (Adler and Newman 2002, p. 69). If advertisers of particular unhealthy products target low-SES populations more than other market groups, it seems not likely due to sinister corporate executives being especially eager to damage those at the bottom of the social ladder. More plausible to us is that such advertising is responsive to reasonably accurate estimates of profit opportunity. As a consequence, the health-related agency of others to whom one is tied in advertising markets influences the advertising one receives. Again, this is not an example of "using" or "utilizing" resources to "garner health advantages," as the advantage is gained without individuals themselves doing anything health-directed at all.

*Habitus.* In consumer theory, if two people buy different quantities of a good, this might be explained by their having different means, but another immediate possibility is that the person who bought more wanted more. The analogy to health is to posit that while everyone might prefer being healthy to being unhealthy, some people may exhibit a stronger and more consistent preference for future good health than others. The idea that differential preferences might have anything to do with health disparities might seem virtually unspeakable in sociological and public health discourses about unequal health outcomes, given how readily it might be construed as "blaming the victim" (Mirowsky and Ross 2003; Klinenberg 2006). Worse, given the current political dominance of narrow neoliberal doctrines about individual sovereignty, such ideas can contribute to discourse that public health advocacy is the meddling of a "nanny state." This, in turn, provides strong incentive toward an explanatory idiom that is predominated by language of "access" to resources and of "constraints," in ways that presume that low-SES individuals share the values of their high-SES counterparts, and differences in outcomes are exclusively the result of agency thwarted.

Indeed, some contend that, as a *matter of definition*, "health disparities reflect unequal opportunities to be healthy" and that "reducing health disparities means giving disadvantaged social groups equal opportunities to be healthy" (Braveman 2006, p. 187). In this light, consider how Adler and Newman (2002, p. 69) discuss education and disparities in health behavior: "Limited education may mean less exposure to information about risk, but the same people may be locked into neighborhoods with poor recreational facilities, fewer stores selling fresh produce, and more advertising for tobacco and alcohol." Smoking and obesity are perhaps today the most prominent SES-related indicators of health behavior, and we have no current evidence that those with lower education are unaware that smoking and obesity are unhealthy (regarding smoking, see Link 2008). Facilities for recreation for low-SES individuals can surely be improved, but it is unknown how much this will reduce the SES gap in exercise. The same can be said for making healthier food more easily available (after all, the fast food outlets frequently lamented in the health inequalities literature offer healthier salads at prices competitive to their burgers). As for advertising, the SES gradient for tobacco use is greater than that for alcohol use, even though tobacco advertising is much more strictly regulated. Emphatically, we agree that equalizing access to health-promoting resources is desirable, but we think sociologists should resist any premature conclusion that SES differences in health are only or even primarily caused by lack of information and "opportunity."

Our goal here is not to draw specific conclusions about SES-based differences in health preferences, but instead to note that, while such differences are not incompatible with the fundamental cause thesis, they do prompt contemplation about the metamechanism(s) responsible. For example, massive differences in the economic quality of life in old age provide one impetus for predicting that those of higher SES might be enduringly more motivated toward maximizing length of life than those of lower SES (Deaton 2002). Similarly, the more people feel in control of their lives and are

spared immediate environmental demands and interpersonal subordination, the more easily they may be able to cultivate a lifestyle of prioritizing long-term health consequences (Mirowsky and Ross 2003). Rose (2007) talks about the rise of a cultural imperative to “live one’s life as a project” with respect to health, and like many cultural developments this may be firstly an elite practice that has only partly diffused down the social hierarchy (see also Aronowitz 2008, p. 7).

“Habitus” is an encompassing term used in some areas of sociology used to refer to basic dispositions of interpretation and action that reflect an actor’s social position (Bourdieu 1984; Sallaz and Zavisca 2007). Differences in habitus regarding health are distinct from either differences in means or spillovers. In our view, some concept like habitus is needed to better integrate theory of SES as a fundamental cause of health with evidence that higher SES individuals better “weave together a healthy lifestyle from otherwise incoherent or diametric practices allocated by subcultural forces” (Mirowsky and Ross 2003, p. 7). Precisely a strength of sociology compared to economics has been its openness to the malleability of preferences to differential experiences and influence, and sociology going forward may be particularly well suited for finding ways of talking about health preferences that move beyond the familiar dichotomy of either asserting a lack of informed opportunity or engaging in blame.

*Institutions.* Both Link and Phelan’s existing work and the two durable narratives discussed above – spillovers and habitus – orient to social institutions as static entities to which individuals may or may not have access and may or may not engage to their health advantage. That is, there is an implicit assumption that agency lies exclusively with the individual, and not with the institutions which may facilitate health gains. Individual-based agency narratives are perpetuated with the assumption that schools, neighborhoods, and physicians provide equitable health returns to all the individuals who come in contact with them, or at least returns that are commensurate with the “resources” put into “harnessing” health benefits. While access is certainly critical, it does not provide the full story for how institutional externalities affect health. An access narrative limits consideration of some of the more sociological aspects of institutions and how they might interact with individual actions and resources to amplify disparities. Therefore, a third durable narrative we see as implicit in existing work is the agentic, dynamic action of institutions.

Using the example of medical care, there is extensive evidence of variation in medical practice according to patient characteristics (including SES, but also gender, race, and age), physician attributes (McKinlay 1996; McKinlay et al. 2002), and healthcare systems (Arber et al. 2004), even when patient case presentation is standardized through the use of vignettes. In previous ethnographic work, we found that a multitude of factors operated from within one healthcare system to further exacerbate SES–health differentials in diabetes care, conditional on patients having access to and utilizing care (Lutfey and Freese 2005). The medical system is not a neutral conduit through which resources are exercised in the way that one might stretch a grocery budget to maximize the purchase of health foods at a store. Rather, it is a dynamic institution that may respond directly to a patient’s efforts to mobilize resources for health, but may also either amplify or mitigate those same efforts. Consider again the example of the business executive who does not make health decisions her top priority. Based on the above studies, we would expect that the well-off business executive and her working class co-worker may receive differential treatment based on SES differences, despite having access to the same healthcare system, insurance, and even physician. Even though the co-worker may actively mobilize her resources to procure the best insurance she can afford, once they are both in the same system, research suggests that they are at risk of being diagnosed and managed differently.

The inclusion of institutional agency adds an important dimension to the fundamental cause story because these dynamics interact differentially with individual SES-related characteristics to affect health. Furthermore, these dynamic externalities change over time. In the case of healthcare, medical diagnosis and treatment vary according to state-of-the-art knowledge of how to mobilize scientific, technological, pharmaceutical, and policy information to improve health outcomes. In this way, the



SES–health link is not simply a matter of whether or not individuals take up public health and medical advice. Access, utilization, and adherence are moot if one’s SES potential for purposive health improvement is undermined by the action of the institution and its agents. Similar dynamics apply for other institutions that have indirect connections to health but robust relationships to both SES and health, including schools, employment (Pager 2007), and the legal system (Massoglia 2008).

## Fundamental Causality and Inquiry

The fundamental cause perspective provides a counterpoint to the dominant epidemiological focus on identifying highly specific risk factors for particular conditions, but its arguments also depend vitally on risk-factor research. The more we know about the causes of disease, the more we can elaborate our understanding of the causes of these causes. For many years in social epidemiology, there has been a disjuncture between highly focused studies of risk factors and studies connecting general health outcomes to broad socioeconomic conditions. The metaphor of “looking upstream” for social causes was the most common framework for thinking about connections between the two (McKinlay 1975). We are now at the point where an array of connections are being made between risk factors and social conditions, and the prospect of approaches that span from “cells to society” or “neurons to neighborhoods” no longer seem like fantastic slogans. For example, Gehlert et al. (2008) outline a series of projects on racial disparities in breast cancer inspired by a “downward causation” model that begins with basic social determinants and proceeds to allostatic load and to environmental mediation of gene expression. While current research often attends to the question of how social inequities get “under the skin,” the fundamental cause perspective calls attention to the concurrent, more encompassing project of understanding how *information* gets under the skin – by emphasizing the centrality of differential returns to knowledge and control *per se* for understanding health disparities.

Conventional risk-factor epidemiology is driven mostly by within-sample comparisons. In a case-control study, ill individuals are matched with healthy controls to try to identify antecedent differences that cause disease. An important contribution of the fundamental cause perspective has been to emphasize the continued importance of quantitative research that is *explicitly comparative* across samples. Comparisons across countries, for example, allow for the possibility of seeing if the magnitudes of health inequalities are linked to broad differences in the distribution of resources or policy regimes (e.g., Beckfield 2004; Olafsdottir 2007; Mackenbach et al. 2008). Likewise, comparisons over time allow for assessment of the effects of changes in dominant threats or available treatments (e.g., Duncan 1996; Lauderdale 2001; Schnittker 2004; Krieger et al. 2008).

To date, research on fundamental causality offers little direct defense against the critique that SES is conceptually too vague and that research on health disparities would be better served by referring simply to the specific indicators that compose SES measures (Deaton 2002; Mirowsky and Ross 2003). There is no evidence for the possibility of a globally applicable SES construct that would allow for equalizing SES by reapportioning its different components, such that  $x$  increase in education would be consistently equivalent to  $y$  increase in income or  $z$  increase in occupational prestige for health outcomes (Warren and Hernandez 2007). Even so, SES remains useful for understanding the macrosociology of disparity and for considering intervention in broad terms. Indeed, a major appeal of the fundamental cause concept is its macrosociological focus, including its potential applicability to places and times without much formal education, where the determinants of social standing may be quite different.

As a complement to macrosocial comparison, ethnographic research allows for the possibility of explicating what fundamental cause relationships actually look like in naturalistic settings. By this, we mean that ethnographic observation affords a unique opportunity to see how the lives of individuals

of differing SES implicate a massive, nonrandom set of circumstances that can be plausibly entertained as contributing, each however slightly, to large ultimate differences in health outcomes. A study of ours was based on observation of two diabetes clinics, one of which served a largely high-SES patient population and the other an overwhelmingly low-SES population (Lutfey and Freese 2005). We focused on diabetes because the strong, well-documented relationship between glucose control and long-term outcomes affords the possibility to observe – even in the relatively brief encounter of the routine clinic visits that we studied – means by which larger trajectories of long-term glucose control are connected to social circumstances (Diabetes Control and Complications Trial 1993). Regardless of the specific fates of individuals observed, ethnographic observation of their experiences and circumstances highlights potential pathways of the aggregate, probabilistic association between SES and diabetes outcomes.

Our findings were a large array of potential mechanisms operating both inside and outside the clinic, as well both internal and external to patients. Additionally, we identified several instances of what we called “compensatory inversions,” in which resources were distributed disproportionately to the patients with the least need for them. For example, the clinic with higher SES patients had far superior diabetes education resources, even though there is ample reason to expect better self-education and self-management from them higher SES patients anyway. Ethnographic observation also allowed us to observe plausibly negative cases of “countervailing mechanisms” that work against higher SES patients. For example, teenage girls and even older women were known to capitalize on the weight loss side effects of uncontrolled diabetes, preferring thinness over appropriate glucose control. Going forward, we hope that comparative quantitative research and in-depth ethnography will complement one another toward the end of providing a fuller picture of the systemic relations between disadvantage and disease.

### **Fundamental Causality and Policy**

Taking seriously the idea that SES has sustained, dynamic influences on health differentials, which transcend the individual-level risk factors commonly identified in public health and epidemiology, poses unique challenges for health policy. Although discussions of fundamental causes of health may be faulted sometimes for being vague about policy implications, one central implication is not at all vague: policies that influence social and economic inequalities are health policies and should be recognized as such. A corollary to this statement is that health disparities will exist so long as there are resource disparities, and so it may be naïve to imagine that the two can be decoupled. Such a conclusion could prompt a figurative throwing up of hands: absent a profound and permanent restructuring of social resources, there may seem no points of leverage for meaningfully reducing disparities.

The emphasis of recent work on the relatively greater importance of education than income at least provides some hope, as substantially reducing education inequalities is not quite so utopic-seeming as doing so for income inequality (Mirowsky and Ross 2003; Mechanic 2007). Beyond a more general equalization of resources, Link and Phelan (2005) discuss the most promising policy implications of a fundamental cause perspective. In summary, their arguments “point to policies that eliminate or reduce the ability to use socioeconomic advantage to gain a health advantage – either by reducing disparities in socioeconomic resources themselves, or by developing interventions that, by their nature, are more equally distributed across SES groups” (Link and Phelan 2005, p. 77). Toward this end, they highlight policies that provide benefit irrespective of individual resources or initiative, as well as policies that attend specifically to the social distribution of knowledge about disease risk and the capacity to act on that knowledge. Although we are supportive of many of the policy ideas they mention, we believe that these ideas also highlight some of the important tensions for sociologists interested in health policy.

1. *Scope of interventions independent of agency.* Two examples of public interventions cited by Link and Phelan that do not depend on voluntary action are “requiring window guards in all high-rise apartments versus advising parents to watch their children carefully” and “banning smoking in public buildings versus advising people to avoid secondhand smoke” (p. 79). Such ideas are a useful riposte, we think, for tendencies toward chronic over-optimism about interventions based purely on providing information. They also, of course, harken to a long tradition of public health triumphs like centralized sanitation and fluoridated water, which brought massive benefits to population health and, for a long while, may have had greater ultimate impact than medical developments (McKeown 1976; McKinlay and McKinlay 1977; Mirowsky and Ross 2003; cf. Timmermans and Haas 2008). Even so, another example of theirs may be especially telling: “air bags rather than seatbeats” to reduce road fatalities. Air bags deploy automatically; seat belts are typically not automatic, and a strong education gradient in seat belt use has been documented (Shinar et al. 2001). Yet, air bags actually work far better in conjunction with seatbelts than they do alone. Moreover, above the minimum standard, there is an array of airbags that can be purchased to further reduce the probability of death in a car crash. As a result, while contextual interventions can be expected to reduce disparities, they still afford opportunities for more effective use by those with the most resources and strongest preferences. While such interventions presumably still reduce disparities by raising the floor of health attainment – e.g., the driver unprotected by air bags – the extent to which that raised floor reduces disparity remains to be empirically assessed. More generally for sociologists interested in inequality and innovation, advancing technologies raise important questions about the social conditions that encourage innovations that improve prospects for the bottom of the health distribution as opposed to further expanding possibilities for those at the top.
2. *Health paternalism.* A main reason for advocating for interventions that minimize the role of individual choice in health is that, for financial and other reasons, lower SES individuals disproportionately make choices that sociologists and many others would rather they did not. At the same time, who pays for those air bags and window guards? If mandating such features is a cost passed on to the consumer by the car manufacturer or landlord, then presumably the people affected most are those who are on the margin of being able to afford a car or apartment. Even if one imagines using taxes to pay for air bags and window guards, one is still proposing a use of money that could be more directly redistributed to lower SES individuals. When we consider reducing agency as a strategy for health disparities, we confront questions both about the morality of restricting choice and about tradeoffs between health and income (Deaton 2002). Our point here is not to take any stance regarding health paternalism ourselves. We do think sociologists interested in policy should be clear that there are no free air bags, and so advocating policies involving mandates also implies thinking about tradeoffs. The issue is especially timely as mandating health insurance premiums has been one of the most controversial aspects of legislation to expand health insurance coverage in the United States. The fundamental causality perspective highlights the tension between the social value placed on individual liberties and the value placed on reducing disparity.
3. *Technology policy and health policy.* Using the example of the high costs of AIDS drugs, Link and Phelan (2005, p. 80) underscore the importance of developing interventions that are broadly accessible and affordable so as to avoid the sorts of cross-national disparities currently observed with those treatments. They also note the importance of constructing interventions that simultaneously address other potential barriers to implementation. As discussed above, we have used the phrase “compensatory inversions” to describe instances in which a health-enhancing resource is distributed disproportionately to higher SES individuals even as lower SES individuals might stand to gain more from them. We think more attention could be directed to ways that compensatory inversions are already rooted in and nourished by the current organization of our health care system. As a major example, the United States health care system makes extensive use of

high-end technological advances, which allows those who can afford it to have some of the most sophisticated treatments available. At the same time, this structure increases costs and so competes with the alternative goal of ensuring the broadest possible access.

At the same time, when considering technological development and health, we think it important for sociology not to view innovation as an innocent or ironic catalyst of disparity. Link (2008) characterizes as an important feature of a "social shaping approach" to health the need "to understand the social distribution of useful knowledge and technology." We agree and believe it important especially to give greater attention to social factors in the development of useful technology. Cross-societal health disparities provide the most transparent examples of the crucial point: consider the difference in the effort for developing treatments for malaria, which relatively few people in wealthier regions get, versus developing treatments for Alzheimer's disease, which relatively few inhabitants of poorer regions get. Private efforts to develop medical innovations are closely related to potential market returns, which in turn is predictably connected to the available resources of the affected (Kremer and Glennerster 2004). For that matter, public efforts can be expected to be associated with the capacity for political influence of particular health constituencies. Even for innovations that already exist and can be produced at relatively low marginal cost (e.g., certain drug treatments), there has been much struggle over solving the social problem of providing them for low cost while they command much higher prices in the United States.

We think medical sociology could participate more in interrogating the development of medical innovations, and also in documenting and understanding this tradeoff and the degree to which it is supported by public health policy and the expenditure of public funds on health research. Much health research defined as "groundbreaking" is directed toward optimizing the health of the optimal patient (Lutfey and Freese 2005). As Link and Phelan (2005, p. 8) put it, "When we create interventions that are expensive and difficult to distribute broadly, we create health disparities." At the same time, there is perhaps often an implicit suggestion of eventual "trickle down" to those with less material and psychological resources for treatment. One can posit that the first step of innovation is figuring out a treatment that can work under relatively ideal conditions (the higher SES condition), and then later work can bring its costs down and facilitate its diffusion (Goldman and Lakdawalla 2005; Glied and Lleras-Muney 2008). Social science has an important role to play in our understanding of how the ultimate health benefits of public expenditures on science are distributed.

4. *Institutional policy leverage.* We urge sociologists interested in health disparities to attend to the institutional settings, medical, and otherwise that mediate SES and health. Knowledge, resources, and interventions are not only distributed at the level of the individual, but also in institutional contexts. To the extent that health policy efforts focus on the former, possible routes for minimizing disparities are truncated at that level. A major contribution sociologists can make to policy efforts is an understanding of how these processes operate at institutional levels and the ways in which framing the problem as one of individual access to "good" schools, physicians, or work settings precludes an evaluation of what happens once people are in those systems. In this sense, a fundamental causes approach calls for more integration of traditional individual-level risk factor interventions with, for example, the Institute of Medicine's (2003) work on quality of care and the role of healthcare providers in contributing to disparities. However, to truly capitalize on existing sociological knowledge of a range of institutions, research on nonhealth institutions should also be included so that we might understand generic dynamics underlying health gradients and how new mechanisms may regenerate in the future. In the global context of health policy research, which predicts challenges such as the disappearance of primary care (McKinlay and Marceau 2008) and expanding pressure for the commercialization of healthcare around the world (Mackintosh and Koivusalo 2005), a policy strategy focused on individual-level interventions may be inadequately prepared to anticipate and address new mechanisms as they emerge and sustain disparities.

## Conclusion

“Fundamental causality” has been one of the most fertile concepts in the recent sociology of health. We have here attempted to provide a systematic exposition of conceptual and theoretical contributions of fundamental causality to the study of SES-based health disparities, focusing especially on the highly influential work of Link and Phelan. We began by articulating fundamental causality as a type of cause, distinct from “distal,” “basic,” “root,” “enduring,” or “important” causes. We next identify four durable narratives for why SES is enduringly related to health. By explicitly developing these narratives, we hope to contribute to moving fundamental causality toward being a more clearly defined theoretical apparatus in medical sociology and health policy. We also address methodological approaches for studying fundamental causality, including quantitative studies of contemporary populations, ethnographic methods, and historical approaches. Finally, we discuss several tensions in sociological thinking about health policy that the fundamental cause concept highlights, drawing again on work from Link and Phelan (2005). In sum, the idea of fundamental causality highlights the importance of placing particularistic studies of risk factors in a larger context of history and inequality, and we anticipate the value of thinking in these terms will be ever more compelling as medical science continues to increase the leverage human beings have over their health.

## References

- Adler NE, Newman K (2002) Socioeconomic disparities in health: pathways and policies. *Health Aff* 21(2):60–76
- Arber S, McKinlay JB, Adams A, Marceau LD, Link CL, O'Donnell AB (2004) Influence of patient characteristics on doctors' questioning and lifestyle advice for Coronary heart disease: a UK/US video experiment. *Br J Gen Pract* 54(506):673–678
- Aronowitz R (2008) Framing disease: an underappreciated mechanism for the social patterning of health. *Soc Sci Med* 67:1–9
- Beckfield J (2004) Does income inequality harm health? New cross-national evidence. *J Soc Behav* 45:231–248
- Bourdieu P (1984) *Distinction: a social critique of the judgment of taste*. Harvard University Press, Cambridge, MA
- Braveman P (2006) Health disparities and health equity: concepts and measurement. *Annu Rev Public Health* 27:167–194
- Christakis NA, Fowler JH (2007) The spread of obesity in a large social network over 32 years. *N Engl J Med* 357:370–379
- Christakis NA, Fowler JH (2008) The collective dynamics of smoking in a large social network. *N Engl J Med* 358:2249–2258
- Cutler DM, Lleras-Muney A (Forth coming) Education and health: evaluating theories and evidence. In: House JS, Schoeni RF, Kaplan G, Pollack H (eds) *The effects of social and economic policy on health*. Russell Sage, New York
- Cutler DM, Deaton A, Lleras-Muney A (2006) The determinants of mortality. *J Econ Perspect* 20(3):97–120
- Deaton A (2002) Policy implications of the gradient of health and wealth. *Health Aff* 21(2):13–29
- Duncan GJ (1996) Income dynamics and health. *Int J Health Serv* 26:419–444
- Gehlert S, Sohmer D, Sacks T, Mininger C, McClintock M, Olopade O (2008) Targeted health disparities: a model linking upstream determinants to downstream interventions. *Health Aff* 27(2):339–349
- Glied S, Lleras-Muney A (2008) Health inequality, education, and medical innovation. *Demography* 45:741–761
- Goldman D, Lakdawalla DN (2005) A theory of health disparities and medical technology. *Contrib Econ Anal Policy* 4(1): Article 8 (<http://www.bepress.com/bejeap>).
- Gottfredson L (2004) Intelligence: is it the epidemiologists' elusive 'fundamental cause' of social class inequalities in health? *J Pers Soc Psychol* 86:174–199
- Graham H (2004) Social determinants and their unequal distribution: clarifying policy understandings. *Milbank Q* 82(1):101–124
- House JS, Kessler RC, Herzog AR, Mero RP, Kinney AM, Breslow MJ (1990) Age, socioeconomic status, and health. *Milbank Q* 68:383–411

- Institute of Medicine (2003) *Unequal treatment: confronting racial and ethnic disparities in healthcare*. The National Academies Press, Washington, DC
- Klinenberg E (2006) *Blaming the Victims: Hearsay, Labeling, and the Hazards of Quick-Hit Disaster Ethnography*. *American Sociological Review* 71:689–698
- Kremer M, Glennerster R (2004) *Strong medicine: creating incentives for pharmaceutical research on neglected diseases*. Princeton University Press, Princeton, NJ
- Krieger N, Rehkopf DH, Chen JT, Waterman PD, Marcelli E, Kennedy M (2008) The fall and rise of US inequities in premature mortality: 1960–2002. *PLoS Med* 5:e46
- Lauderdale D (2001) Education and survival: birth cohort, period, and age effects. *Demography* 38:551–561
- Lieberman S (1985) *Making it count: the improvement of social research and theory*. University of California Press, Berkeley, CA
- Link BG (2008) Epidemiological sociology and the social shaping of population health. *J Health Soc Behav* 49:367–384
- Link BG, Phelan JC (1995) Social conditions and fundamental causes of disease. *J Health Soc Behav Extra Issue*:80–94
- Link BG, Phelan JC (1996) Understanding sociodemographic differences in health –The role off fundamental social causes. *Am J Public Health* 86:471–473
- Link BG, Phelan JC (2000) Evaluating the fundamental cause explanation for social disparities in health. In: Bird CE, Conrad P, Fremont AM (eds) *Handbook of medical sociology*. Prentice Hall, Englewood Cliffs, NJ
- Link BG, Phelan JC (2002) McKeown and the idea that social conditions are the fundamental causes of disease. *Am J Public Health* 92(5):730–732
- Link BG, Phelan JC (2005) Fundamental sources of health inequalities. In: Mechanic D, Rogut LB, Colby DC, Knickman JR (eds) *Policy challenges in modern health care*. Rutgers University Press, New Brunswick, NJ, pp 71–84
- Link BG, Northridge ME, Phelan JC, Ganz M (1998) Social epidemiology and the fundamental cause concept: on the structuring of effective cancer screens by socioeconomic status. *Milbank Q* 76(3):375–402
- Link BG, Phelan Jo C, Miech R, Westin EL (2008) The resources that matter: fundamental social causes of health disparities and the challenge of intelligence. *J Health Soc Behav* 49:72–91
- Lutfey K, Freese J (2005) Toward some fundamentals of fundamental causality: socioeconomic status and health in the routine clinic visit for diabetes. *Am J Sociol* 110(5):1326–1372
- Mackenbach JP, Stirbu I, Roskam AJ, Schaap MM, Menvielle G, Leinsalu M, Kunst AE (2008) Socioeconomic inequalities in health in 22 European countries. *N Engl J Med* 358:2468–2481
- Mackintosh M, Koivusalo M (eds) (2005) *Commercialization of health care: global and local dynamics and policy responses*. [Series on social policy in a development contextg, United Nations Research Institute for Social Development]. Macmillan, New York
- Marmot M (2004) *The status syndrome: how social standing affects our health and longevity*. Holt, New York
- Massoglia M (2008) Incarceration as exposure: the prison, infectious disease, and other stress-related illness. *J Health Soc Behav* 49(1):56–71
- McKeown T (1976) *The Role of Medicine: Dream, Mirage, or Nemesis*. Nuffield Provincial Hospitals Trust, London
- McKinlay JB (1975) A case for refocusing upstream: the political economy of illness. In: Enelow AJ, Henderson JB (eds) *Behavioral aspects of prevention*. American Heart Association, Dallas, TX, pp 9–25
- McKinlay JB (1996) Some contributions from the social system to gender inequalities in heart disease. *J Health Soc Behav* 37(1):1–26
- McKinlay JB, Marceau LD (2008) When there is no doctor: reasons for the disappearance of primary care physicians in the U.S. during the early 21<sup>st</sup> century. *Soc Sci Med* 67:1481–1491
- McKinlay JB, McKinlay SM (1977) The questionable contribution of medical measures to the decline of mortality in the United States in the twentieth century. *Milbank Q* 55:405–428
- McKinlay JB, Lin T, Freund K, Moskowitz M (2002) The unexpected influence of physician attributes on clinical decisions: results of an experiment. *J Health Soc Behav* 43(1):92–106
- Meara ER, Richards S, Cutler DM (2008) The gap gets bigger: changes in mortality and life expectancy, by education, 1981–2000. *Health Aff* 27(2):350–360
- Mechanic D (2007) Population health: challenges for science and society. *Milbank Q* 85(3):533–559
- Mirowsky J, Ross CE (2003) *Education, social status, and health*. Aldine de Gruyter, New York
- Olafsdottir S (2007) Fundamental causes of health disparities: stratification, the welfare state, and health in the United States and Iceland. *J Health Soc Behav* 48:239–253
- Pager D (2007) *Marked: race, crime, and finding work in an era of mass incarceration*. University of Chicago Press, Chicago, IL
- Phelan JC, Link B (2005) Controlling disease and creating disparities: a fundamental cause perspective. *J Gerontol* 60B(Special Issue II):27–33

- Phelan JC, Link B, Diez-Roux A, Kawachi I, Levin B (2004) Fundamental causes of social inequalities in mortality: a test of the theory. *J Health Soc Behav* 45:265–285
- Preston SH, Elo IT (1995) Are educational differences in adult mortality increasing in the United States? *J Aging Health* 7:476–496
- Rose N (2007) *The politics of life itself: biomedicine, power, and subjectivity in the twenty-first century*. Princeton University Press, Princeton, NJ
- Sallaz JJ, Zavisca J (2007) Bourdieu in American sociology, 1980–2004. *Annu Rev Sociol* 33:21–41
- Sankar P, Cho MK, Condit CM, Hunt LM, Koenig B, Marshall P, Lee SS, Spicer P (2004) Genetic research and health disparities. *JAMA* 291(24):2985–2989
- Schnittker J (2004) Education and the changing shape of the income gradient in health. *J Health Soc Behav* 45:286–305
- Shinar D, Schechtmana E, Compton R (2001) Self-reports of safe driving behaviors in relationship to sex, age, education and income in the U.S. adult driving population. *Accid Anal Prev* 33(1):111–116
- Smith JP (2007) The impact of socioeconomic status on health over the life-course. *J Hum Resour* 42(4):739–764
- Steenland K, Hu S, Walker J (2004) All-cause and cause-specific mortality by socioeconomic status among employed persons in 27 U.S. states, 1984–1997. *Am J Public Health* 94:1037–1042
- The Diabetes Control and Complications Trial Research Group (1993) The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med* 329:977–986
- Timmermans S, Haas S (2008) Toward a sociology of disease. *Sociol Health Illn* 30:659–676
- Warren JR, Hernandez EM (2007) Did socioeconomic inequalities in morbidity and mortality change in the United States over the course of the twentieth century? *J Health Soc Behav* 48:335–351