Volume Title: Research Findings in the Economics of Aging

Volume Author/Editor: David A. Wise, editor

Volume Publisher: The University of Chicago Press

Volume ISBN: 0-226-90306-0

Volume URL: http://www.nber.org/books/wise08-1

Conference Dates: May 10-13, 2007

Publication Date: February 2010

Chapter Title: Social Interactions and Smoking

Chapter Author: David M. Cutler, Edward L. Glaeser

Chapter URL: http://www.nber.org/chapters/c8196

Chapter pages in book: (123 - 141)

5

Social Interactions and Smoking

David M. Cutler and Edward L. Glaeser

5.1 Introduction

A large and growing literature suggests that individual choices are influenced by the choices of their friends and neighbors. These peer effects have been found in dropping out, unemployment, crime, pregnancy, and many other settings (Crane 1991; Case and Katz 1991; Glaeser Sacerdote, and Scheinkman 1996; Topa 2001; Brock and Durlauf 2001; Kuziemko 2006). The older work in this literature was criticized because the company you keep is rarely random (Manski 1993). Newer work in this area has documented peer effects in settings where there is real random assignment, such as college dormitories (Sacerdote 2001).

There are many reasons to think that peers matter for health-related behaviors. In many cases, health-related behaviors are more fun to do when others are doing them too (drinking, for example). Peers are also a source of information (the benefits of a mammogram) or about what is acceptable in society (the approbation accorded smokers). A recent study suggested that a good part of the obesity "epidemic" in the United States is spread from person to person, in a manner reminiscent of viral infections (Christakis and Fowler 2007).

These interpersonal complementarities can have enormous social impact. In addition to helping us understand how health behaviors operate, they magnify the impact of policy interventions. The existence of social inter-

David M. Cutler is the Otto Eckstein Professor of Applied Economics at Harvard University and a research associate of the National Bureau of Economic Research. Edward L. Glaeser is the Fred and Eleanor Glimp Professor of Economics at Harvard University and a research associate of the National Bureau of Economic Research.

We are grateful to Alice Chen for research assistance, to Arie Kapteyn for comments, and to the National Institutes on Aging (Grant# P01 AG005842) for research support.

actions implies that a policy intervention has both a direct effect on the impacted individual and an indirect, as that person's behavior impacts everyone around. These indirect effects create a social multiplier where the predicted impact of interventions will be greater when the interventions occur at large geographic levels than when they occur individually (Glaeser, Sacerdote, and Scheinkman 2003). The social multiplier also suggests that parameter estimates from aggregate regressions can mislead us about individual level parameters.

In this chapter, we assess the evidence on social interactions in one particularly important health-related behavior: smoking. There are a number of reasons we might expect to see social interactions in smoking, as we discuss in section 5.2. These include direct social interactions (where one person's utility is affected by whether others are doing the same thing); the social formation of beliefs; and supply-side interactions from market creation in a situation in fixed costs.

Section 5.3 lays out the empirical implications of social interactions. The most straightforward implication of social interactions is that an exogenous variable that increases the costs of a behavior for one person will decrease the prevalence of that behavior is his or her peers. Social interactions models also predict excess variance in smoking rates across aggregates. Finally, the existence of social interactions implies that the measured impact of an exogenous variable on an outcome becomes larger at higher levels of aggregation.

In sections 5.3 and 5.4, we look at these three empirical predictions. At the individual level, we examine the impact of workplace smoking bans on spousal smoking. Evans, Farrelly, and Montgomery (1999) show that workplace bans have a significant impact on the probability that an individual will smoke and that these bans survive various estimation strategies that address selection of smokers into smoke-friendly workplaces. We look at whether people are more likely to smoke if their spouse smokes, using workplace smoking bans as an instrument for spousal smoking. The independent variable (IV) estimate is large: we estimate that an individual whose spouse smokes is 40 percent more likely to smoke. The instrumental variables estimate is higher for men than for women, suggesting that men are more influenced by spousal smoking. These effects are also stronger for people with some college than for people with college degrees or people who were high school dropouts.

In section 5.5, we turn to the other empirical implications of social interactions. We first show that the impact of smoking bans appears to be greater at the area level than at the individual level. At an individual level, a workplace ban reduces the probability of smoking by about 5 percent. At the metropolitan area level, a 10 percent increase in the share of workers facing workplace bans reduces the share of people who smoke by more than 3 percent—six times greater than the .5 percent predicted by the individual model. At the state level, the social multiplier rises to more than ten.

We also examine the prediction that social interactions create excess variance of aggregate smoking rates. We find that the standard deviation of smoking rates across metropolitan areas or states are about seven times higher than the rates that would be predicted if there were no social interactions and if there were no exogenous variables that differed across space. Since there are significant exogenous variables that differ across space, we do not put complete stock in these numbers. Still, these high variances provide evidence supporting the existence of social interactions in smoking.

Section 5.6 turns to the question of whether social interactions can help us make sense of the time series of smoking. Social interactions predict *s*-shaped adoption curves and changes are a function of current levels of smoking. A simple regression suggests that social interactions are not obvious in the national dynamics of cigarette prevalence, but our samples for this regression are small. Section 5.7 concludes.

5.2 Sources of Social Interactions

Why should one person's smoking increase his neighbor's tendency to smoke? There are three broad categories of reasons for such social interactions: (a) direct social interactions, including social approval and stigma; (b) the social formation of beliefs, and (c) market-mediated spillovers that occur because of fixed costs in the provision of healthy or unhealthy behavior. In this section, we briefly review these three possible reasons for interpersonal complementarities in smoking and other health-related behaviors.

The first reason that one person's smoking, or eating or exercise, might positively influence a neighbor's choices is that it is more pleasant to do something together than alone. This is most obvious in the context of eating, where it is more pleasurable (most of the time) to eat with others rather than eating alone. Because of the desire to eat together, people are more likely to go to donut shops, steak houses, or McDonald's, if their friends are also doing so. Drinking is also a social activity; if one's friends like to drink in bars, the returns from going to bars rises. Smoking and exercise may be somewhat less social activities, but many people like to exercise or smoke with friends around.

Conversely, smoking around a nonsmoker can be much less pleasant because of the discomfort caused by secondhand smoke to a nonsmoker. While there may be debate about the health consequences of secondhand smoke, there is less disagreement about whether nonsmokers dislike smoke. If a smoker has some degree of altruism for the uncomfortable nonsmoker, or if the nonsmoker chooses to reciprocate his discomfort by scolding the smoker, then this will decrease the returns to smoking around nonsmokers.

A second reason for social interactions in health behaviors is that beliefs may themselves be formed through social learning. One type of social learning model suggests that people infer truth from the behavior of others (e.g., Ellison and Fudenberg 1993). A person may not know whether moderate drinking is good or bad, but they can get guidance on this by watching others they believe have more information. In these models, the presence of friends and neighbors who smoke, drink, or exercise will provide evidence about the benefits of these activities. Conversely, the absence of smoking will be taken to mean that there is something wrong with lighting up.

Of course, conversation also transmits information (e.g., DeMarzo, Vayanos, and Zweibel 2003). If smoking, or any other harmful activity, increases one's belief in the net benefits of that activity—perhaps because of cognitive dissonance—then smokers are likely to articulate the view that cigarettes are pleasurable or not harmful. These views will then be transmitted in conversation and perhaps persuade some peers that smoking is less harmful. The power of these views will depend, of course, on the extent to which other messages about the benefits or harms of the activity are being regularly broadcast.

The third reason for social interactions works through the market. The typical assumption about markets is that supply curves slope up: when more people consume a good, the price of that good rises. This creates a negative social interaction; more people smoking will drive up the price of cigarettes, and discourage some marginal smokers from smoking. However, as George and Waldfogel (2003, 2006) have recently emphasized, in the presence of fixed costs these negative market-based social interactions can be reversed. Suppliers are only likely to pay the fixed costs to set up if the market size is sufficiently high. In that case, the market creates a strong positive social interaction.

This market-based interpersonal complementarity is more likely in goods with fixed costs, such as restaurants, grocery stores, bars, or health clubs. Cigarettes production itself has large fixed costs, but since transport costs are low, cigarette availability does not depend on local market size. However, several studies have shown that healthy foods are hard to buy in low income areas, presumably because of limited demand. The presence of health clubs and bars also depend on the presence of sizable local demand.

The relative importance of these different types of social interactions will differ across behaviors. Direct interactions and belief formation seem more important for smoking. Market-based interactions are more likely to be important for exercise and consumption of healthy food. In the next section, we will not distinguish between these different sources of social interactions but discuss more generally the empirical implications of interpersonal complementarities in health-related behaviors.

5.3 Empirical Tests of Social Interactions

The literature on social interactions has broadly identified four different empirical implications of social interactions. First, social interactions imply that a person is more likely to undertake an activity when his or her peers are also undertaking that activity. Second, the existence of social interactions implies a social multiplier, where the impact of some exogenous characteristic on the outcome at an individual level is much smaller than the impact of that same characteristic on the outcome at an aggregate level. Third, social interactions imply high levels of variance in the activity across space (Glaeser, Sacerdote, and Scheinkman 1996). Fourth, in a dynamic setting, social interactions lead to an *S*-shaped adoption curve. In this section, we present a particularly simple social interaction model that illustrates the first three points. In section 5.6, we discuss a dynamic model.

We start with a simple model of social interactions. We assume that individual i receives private benefits from an activity, X_i, of A_iX_i, where A_i differs across individuals. The cost of the activity is $.5X^2$. To capture social interactions, we assume that benefits increase by b times that average consumption of X among person i's friends, which we denote $\hat{X}_i$. The utility of individual i is therefore $(A_i + b\hat{X}_i)X_i - X_i^2$. When individuals set marginal benefits equal to marginal costs, the optimal level of X will satisfy $X_i = A_i + b\hat{X}_i$.

Aggregating this relationship implies that $\hat{X}_i = \hat{A}_i/(1-b)$, where $\hat{A}_i$ refers to the average value of A in i's peer group. Substituting this term in implies that individual X will equal $A_i + b\hat{A}_i/(1-b)$. If b is greater than 1/2, then the impact of average "*A*" is greater than the impact of individual "*A*".

These calculations deliver the basic empirical implications of social interactions models. First, there will be greater variation in the outcome across space than would be predicted based on individual differences alone. Within groups, the variance of the outcome will be $Var(A_i)$ while the variance of outcomes across groups will equal $Var(\hat{A}_i)/(1-b)^2$. If there are N people in each group who are allocated randomly, then $Var(\hat{A}_i) = Var(A_i)/N$, so in that case, the ratio of the aggregate variance to the individual within group variance should equal $1/[N(1-b)^2]$. High group level variance is a sign that "b" is high.[1]

While we implement this test, we note one obvious difficulty with it: the ratio of across to within group variance is likely to be biased upwards because of omitted characteristics that differ at the group level. For example, if exogenous tastes for smoking differ across areas and we cannot control for tastes, we will attribute the variation in smoking rates across areas to social spillovers rather than tastes. One method of dealing with this problem is to control extensively for observable characteristics and then to assume that the heterogeneity across groups in the unobservable characteristics is some multiple of the heterogeneity across groups in observable characteristics.

A second implication of the model is the existence of a social multiplier.

1. We conduct our test using standard deviations: the ratio of the standard deviation at the group level, to the standard deviation at the individual level divided by the square root of N is an estimate of $1/(1-b)$.

To see this, assume that $A_i = a_i + \delta z_i$ where δ is a constant and z_i is an exogenous characteristic such as income or public policy regulations. In this case, regressing the outcome on z at the individual level will give a coefficient of δ, while the same regression at the aggregate level will give a coefficient of $\delta/(1 - b)$. Thus, the group level relationship will be stronger than individual relationship, which is the definition of a social multiplier.

The most common empirical approach to social interactions has been at the individual level, estimating a regression of one person's outcomes on the outcomes of a neighbor. The reflection problem (Manski 1993) means that a direct regression of this sort does not recover the parameter b. For example, assume a peer group of two people, i and j. Then, person i's outcome is $A_i + bX_j$ and person j's outcome is $A_j + bX_j$. Solving these two equations implies that person i's outcome equals $(A_i + bA_j)/(1 - b^2)$ and person j's outcome equals $(A_j + bA_i)/(1 - b^2)$. Straightforward analysis shows that a univariate regression where person i's outcome is regressed on person j's outcome does not yield the parameter b, but rather $2b/(1 + b^2)$.

External factors can help us with this problem, however. Specifically, if $A_i = a_i + \delta z_i$ and z_j is used as an instrument for A_j then the instrumental variables estimate of the social interaction $(\mathrm{Cov}(A_i,z_j)/\mathrm{Cov}(A_j,z_j))$ will equal b. We will follow this approach in our analysis.

5.4 Social Interactions in Smoking: Direct Tests

Surely a spouse is among the most important of all social influences. For all of the reasons previously discussed, we would expect the influence of behaviors to be particularly large within a family. In addition, smoking might be sensitive to peers or other people similarly situated. In this section, we look at the influence of one spouse's smoking decisions on the smoking propensity of the other spouse. We also look at the influence of smoking rates for people with similar demographic characteristics. Clearly the decision of two married people or friends to smoke is endogenous. To address the endogeneity issues just discussed, we follow Evans, Farrelly, and Montgomery (1999) and use the presence of workplace smoking bans as an instrument for the smoking of one spouse.

We use the Current Population Survey (CPS) tobacco supplement data for information on smoking rates and workplace smoking bans. The CPS asks about smoking and smoking bans in four periods: 1992 and 1993, 1995, 1998, and 2002. We sample people between the ages of fifteen and sixty-four. The smoking data is asked of everyone. The smoking ban question is asked only of indoor workers. We discuss this more in the following.

Table 5.1 shows the means and standard deviations from this data source. Between 1992 and 2002, the overall smoking rate declined from 25 percent to 20 percent, a reduction of one-fifth. The decline for indoor workers, who

Table 5.1 **Trends in smoking rates and smoking bans (%)**

Measure	1992–1993	1995	1998	2002
Smoking rate, overall	25	25	24	20
Smoking rate, indoor workers	24	24	23	20
Percent with smoking ban, overall	35	42	44	45
Percent with smoking ban, indoor workers	66	75	78	79

Note: The sample is self-respondents aged fifteen to sixty-four from the Current Population Survey. Data are weighted using sample weights.

are those effected by smoking bans, was similar: 24 percent in 1992 and 1993 to 20 percent in 2002.

Smoking bans for indoor workers were spreading rapidly in the 1990s. While the overall share of the sample with a smoking ban increases from 35 percent in 1992 and 1993 to 45 percent in 2002, the share of the indoor workers with smoking bans increased from 66 percent in 1992 and 1993 to 79 percent ten years later. The current omnipresence of workplace bans represents a remarkable change over twenty-five years. Evans, Farrelly, and Montgomery (1999) report that as late as 1985, only one-quarter of workplaces banned smoking.

As Evans, Farrelly, and Montgomery (1999) discuss, the estimated impact of smoking bans on smoking may be biased because of sorting across jobs. Smokers may choose jobs that are particularly smoke-friendly, and this will cause a negative correlation between workplace bans and smoking that does not reflect the impact of the bans. Their own instrumentation strategy suggests that this selection (within indoor jobs) is relatively weak. We have no comparable sources of exogenous variation. As such, we will look at the impact of workplace bans directly without using instruments.

We start by looking at the impact of smoking bans on the smoking rates of people affected by them. To do this, we estimate a model of smoking rates as a function of demographics and the presence of a smoking ban:

$$\text{Smoke}_i = \beta_0 + \beta_1 \cdot \text{Smoking Ban}_i + Z_i\beta + \varepsilon_i, \tag{1}$$

where i denotes individuals and Z is the control variables. We include a number of standard controls: age and its square, gender, family size, family income, a dummy for missing income, education ($<$ high school, high school, some college, college grad, $>$ college), race/ethnicity (white, black, Hispanic, other race), marital status (married, divorced, separated, widowed, never married), industry dummies, occupation dummies, a dummy for whether the person is employed, and a dummy for whether the person is an indoor worker. We also control for metropolitan area and year fixed effects so that our results reflect changes in smoking bans within regions over time.

The first column in table 5.2 shows our basis results. Since the dependent variable is dichotomous, we report marginal effects from a Probit regression. We estimate that workers who face workplace smoking bans are 4.6 percent less likely to be smokers. The coefficient is highly statistically significant. The magnitude here is similar to that found in Evans, Farrelly, and Montgomery (1999), who estimated that smoking bans reduce workplace smoking by 5 percent.

We are less concerned with the other variables, but some are worthy of note. Surprisingly, we do not find a significant effect of cigarette taxes on smoking. The coefficient is negative, as expected, but not statistically significant. It may be that by the late 1990s, the most price sensitive smokers have already left the market. More education is negatively related to smoking, with large coefficients. College graduates are 15 percent less likely to smoke than high school graduates. Blacks and Hispanics are less likely to smoke than are whites, and employed people smoke less.

We now turn to the models including spillovers. In regression (2), we show the ordinary least squares regression when individual smoking is regressed on all of the variables in the first regression and on an indicator variable for whether the spouse smokes.[2] The regression shows that people whose spouse smokes are 21 percent more likely to smoke themselves. We would normally expect this coefficient to be biased upwards both because of the endogeneity of spousal smoking and because of selection of spouses.

Regression (3) looks at the spillovers of smoking in a more general peer group. As is common in the literature, we define the peer group as people in the same metropolitan area and cohort group within the same metropolitan area and with the same age (fourteen to thirty, thirty-one to fifty, and fifty-one to sixty-four) and education level ($<$ high school, high school, come college, college graduate). There is a very high correlation of smoking rates across people in a common reference group. The coefficient on reference group smoking is 0.8, which means that as the share of peers that smokes increases by 10 percent, the probability that an individual will himself smoke increases by 8 percent. As in the case of the spousal smoking coefficient, we expect this coefficient to be biased upwards because individuals influence their peers and because of omitted variables that are correlated across peers.

The obvious solution in each case is instrumental variables. In the case of spousal smoking, we instrument with whether the spouse has a smoking ban at work. In the case of peer group smoking, we instrument with the share of the peer group that has a smoking ban at work. Regressions (4) and (5) show these results—the former for spousal smoking only, and the latter for spousal and reference group smoking.

2. Since this is a prelude to the instrumental variables estimates, we also include dummies for whether the spouse is employed, and whether the spouse is an indoor worker.

Table 5.2 **Explaining smoking decisions**

	Individual ban only	With peer effects			
Independent variable	OLS (1)	OLS (2)	OLS (3)	IV (4)	IV (5)
Smoking					
Smoking ban	–0.046	–0.043	–0.042	–0.041	–0.041
	(0.005)***	(0.005)***	(0.005)***	(0.005)***	(0.005)***
Spouse smokes	—	0.211	0.180	0.401	0.400
		(0.005)***	(0.006)***	(0.082)***	(0.084)***
Reference group smoking rate	—	—	0.880	—	0.050
			(0.012)***		(0.285)
Cigarette tax	–0.005	–0.006	0.006	–0.006	–0.005
	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)
Demographics					
Age	0.025	0.024	0.013	0.023	0.023
	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.004)***
Age^2	–0.0003	–0.0003	–0.0002	–0.0003	–0.0003
	(9.4E-6)***	(1.1E-5)***	(1.1E-5)***	(1.2E-5)***	(4.5E-5)***
Female	–0.036	–0.04	–0.039	–0.044	–0.044
	(0.003)***	(0.003)***	(0.003)***	(0.003)***	(0.004)***
Family size	–0.018	–0.017	–0.016	–0.017	–0.017
	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Ln(family inc)	–0.047	–0.044	–0.038	–0.041	–0.041
	(0.002)***	(0.002)***	(0.002)***	(0.003)***	(0.004)***
Income missing	–0.524	–0.487	–0.421	–0.458	–0.455
	(0.024)***	(0.026)***	(0.026)***	(0.030)***	(0.038)***
< High school	0.019	0.017	0.016	0.016	0.014
	(0.006)***	(0.006)***	(0.005)***	(0.006)***	(0.006)**
Some college	–0.05	–0.045	0.015	–0.041	–0.036
	(0.004)***	(0.004)***	(0.004)***	(0.004)***	(0.020)
College grad	–0.148	–0.137	0.034	–0.127	–0.114
	(0.005)***	(0.005)***	(0.005)***	(0.006)***	(0.056)**
> College	–0.17	–0.156	0.014	–0.143	–0.13
	(0.005)***	(0.005)***	(0.006)**	(0.008)***	(0.055)**
Black	–0.078	–0.073	–0.067	–0.069	–0.069
	(0.005)***	(0.005)***	(0.005)***	(0.005)***	(0.006)***
Hispanic	–0.13	–0.122	–0.096	–0.116	–0.114
	(0.005)***	(0.005)***	(0.005)***	(0.006)***	(0.011)***
Other race	–0.056	–0.052	–0.051	–0.049	–0.049
	(0.007)***	(0.006)***	(0.006)***	(0.006)***	(0.007)***
Divorced	0.098	0.125	0.113	0.154	0.153
	(0.005)***	(0.006)***	(0.006)***	(0.014)***	(0.014)***
Separated	0.108	0.135	0.122	0.165	0.164
	(0.010)***	(0.010)***	(0.011)***	(0.017)***	(0.017)***
Widowed	0.066	0.093	0.083	0.122	0.122
	(0.010)***	(0.012)***	(0.011)***	(0.016)***	(0.017)***
Never married	0.03	0.055	0.048	0.082	0.082
	(0.004)***	(0.005)***	(0.005)***	(0.012)***	(0.013)***

(*continued*)

Table 5.2 (continued)

	Individual ban only	With peer effects			
Independent variable	OLS (1)	OLS (2)	OLS (3)	IV (4)	IV (5)
Employed	–0.074 (0.008)***	–0.071 (0.008)***	–0.059 (0.008)***	–0.068 (0.008)***	–0.068 (0.008)***
Indoor worker	0.041 (0.006)***	0.038 (0.006)***	0.037 (0.006)***	0.035 (0.006)***	0.036 (0.007)***
Spouse employed	—	–0.009 (0.005)**	–0.008 (0.004)	–0.012 (0.005)**	–0.012 (0.005)**
Spouse indoor worker	—	–0.009 (0.004)**	–0.005 (0.003)	–0.014 (0.004)***	–0.014 (0.005)***
Percent reference group employed	—	—	–0.074 (0.013)***	—	0.004 (0.030)
Percent reference group indoor worker	—	—	–0.03 (0.011)***	—	–0.031 (0.012)**
MSA dummy variables	Yes	Yes	Yes	Yes	Yes
Year dummy variables	Yes	Yes	Yes	Yes	Yes
N	195,579	195,579	195,579	195,579	195,579
R^2	0.10	0.11	0.17	0.10	0.11

Notes: Data are from CPS Sept. 1992/May 1993, Sept. 1995, Sept. 1998, and Feb. 2002 Tobacco Supplement Surveys. Sample composition is of people aged fifteen to sixty-four. All regressions also include major industry (twenty-one dummies) and major occupation (thirteen dummies) effects, and are weighted by the self-response supplement sample weight. Models for individuals and spouses are clustered by family id. Models including cohort effects are clustered by the MSA cohort education level with cohort ages of fourteen to thirty, thirty-one to fifty, and fifty-one to sixty-four and education levels of less than high school, high school, some college, and college graduates or higher. Spouse smokes instrumented by spouse smoking ban, and reference group smoking rate instrumented by share of reference group with a smoking ban. OLS = ordinary least squares; IV = independent variable. Dashed Cells = not included in regression.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

The instrumentation has very different effects on the estimated spouse and reference group coefficients. When we instrument using smoking bans facing one's spouse, we find that the estimated impact of spousal smoking increases to .4, so that people whose spouses smoke are 40 percent more likely to smoke themselves. While the magnitude of this coefficient is not unreasonable, we are somewhat skeptical about the fact that the estimated coefficient rises. One interpretation of this might be that we are not measuring the intensity of spousal smoking, and working in a place without a ban might be particularly correlated with intensive smoking. An alternative interpretation is that spouse's workplace smoking bans are correlated with other characteristics, like the prosmoking atmosphere in one's social group, that we cannot adequately control for.

In regression (5), we see that the instrumental variables approach completely eliminates the estimated impact of peer smoking on an individual's decision to smoke. While the standard error is large (29 percent), the coefficient is very small (5 percent). The coefficient on spousal smoking, in contrast, is essentially unchanged. One interpretation of these results is that spousal smoking does have spillovers, but peer group smoking does not. Another view is that our instrumental variables peer group coefficient is not precisely estimated enough to really say much about the impact of peers on smoking.

One question commonly speculated about is how spillovers differ by demographic group. One often hears that less educated groups might be more responsive to peer influences, though information dissemination is perhaps greater in better educated groups. In table 5.3, we estimate the spillover effects separately for different population subgroups. The regressions are all similar to those in table 5.2, though we only report the coefficients on

Table 5.3 **Examining the response to smoking bans by demographic group (instrumental variable estimates)**

Group	Smoking ban	Spouse smokes	Reference group smoking rate	N	R^2
All	–0.041	0.400	0.050	195,579	0.11
	(0.005)***	(0.084)***	(0.285)		
By gender					
Men	–0.052	0.502	–0.002	86,321	0.1
	(0.008)***	(0.196)**	(0.416)		
Women	–0.029	0.365	–0.264	109,258	0.04
	(0.006)***	(0.073)***	(0.628)		
By education					
< High school	–0.033	–0.0080	–0.054	29,392	0.18
	(0.014)**	(0.525)	(2.235)		
High school	–0.050	0.289	–3.198	61,744	—
	(0.012)***	(0.261)	(5.203)		
Some college	–0.042	0.663	–0.269	52,175	—
	(0.011)***	(0.177)***	(0.668)		
College +	–0.020	0.346	1.201	52,268	0.07
	(0.008)***	(0.191)	(1.148)		

Notes: The reference group is based on the MSA cohort education level. All regressions include age, age squared, family size, log(family income), missing income dummy, three indicators for ethnicity, four indicators for marital status, cigarette tax (state + federal), twenty-one industry indicators, and thirteen occupation indicators. Regression for all, men, and women also include four indicators for educational attainment. Regressions for all and education bins include indicator for gender. Spouse smokes instrumented by spouse smoking ban, and reference group smoking rate instrumented by share of reference group with a smoking ban. Regressions weighted by self-response supplement weight.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

workplace bans, spousal smoking, and peer group smoking. The first row in the table reports our benchmark results from column (5) of table 5.2.

The next two rows report these results separately for men and women. Workplace smoking bans have a larger impact on men (5.2 percent) than on women (2.9 percent). This may be because men are more likely to work full time, or because men infer more from a workplace smoking ban than do women. Men are also more sensitive to spousal smoking than are women. The coefficient on (instrumented) spousal smoking is 0.50 for men and 0.37 for women. According to these findings, wives have a bigger impact on husbands than husbands have on wives. The reference group smoking rate is insignificant for both genders.

The next four rows show the results for four separate education groups: high school dropouts, high school graduates with no college, people with some college education, and people with college degrees. The impact on workplace bans is strongest for those individuals in the middle education categories. The impact of spousal smoking is strongest for people with some college and weakest for people who are high school dropouts. The reference group effects differ substantially across education subgroups but are never statistically significant.

Overall, these findings support the idea of a substantial social interaction in smoking between spouses. While we are not confident that the right coefficient is .4, rather than .2, we are reassured by the fact that the positive social spillover is robust to our instrumental variables strategy. The reference group may also be important, but the fact that it is not robust to our instrumental variables strategy makes us less confident about its strength.

5.5 Social Multipliers and Excess Variance in Smoking

We now turn to other evidence for social spillovers in smoking: variability across groups and social multipliers. We start with nonparametric evidence: the variability in smoking rates across groups. At the individual, our estimated smoking rate of 24 percent implies a standard deviation of .43. If there were no omitted variables across metropolitan areas and if there were no social interactions, then this variance should decline substantially with group size. Specifically, the standard deviation of smoking rates across a group of size N should equal $.42/\sqrt{N}$.

Our metropolitan area samples have, on average, 3,238 individuals, which implies that the standard deviation of smoking rates across groups should equal approximately .008. As table 5.4 shows, this is approximately one-sixth of the actual variation in smoking rates across our metropolitan area samples. At the state level, our average sample size is 10,684, which implies that the standard deviation of smoking rates across state groups should equal approximately .004. Again, the actual standard deviation is almost seven times larger than this amount.

Table 5.4 **The variability of smoking across areas**

	Average observations per unit	Predicted standard deviation	Actual standard deviation	Ratio: Actual/predicted
Individual	1	0.427	0.427	—
MSA	3,238	0.008	0.046	6.1
State	10,684	0.004	0.027	6.5

Note: The sample is self-respondents aged fifteen to sixty-four from the Current Population Survey. Data are weighted using sample weights. Dashed cell = ratio not appropriate.

Using the calculations in section 5.2, an aggregate to individual standard deviation of 6 suggests a value of b of .83. Surely, this estimate is biased upwards because of omitted group level characteristics. Nonetheless, there is a high level of variation at the group level, which supports the idea that social interactions may be important in smoking.

A third test for social multipliers is to look at the impact of external factors on smoking rates at the individual and group level. As section 5.2 pointed out, in a situation of social multipliers, the aggregate impact of a particular factor will be greater than the individual impact. We test this using the individual, metropolitan statistical area (MSA), and state-level samples. The basic approach of these regressions is to regress smoking on the same characteristics at the individual, metropolitan area and state level. If social interactions are important then we should expect the impact of characteristics to become more important at higher levels of aggregation (Glaeser, Sacerdote, and Scheinkman 2003).

In principle, a social multiplier could show up in any variable, but we would be less inclined to see it in variables that are strongly correlated with social groupings. For example, even though age is correlated with smoking, we might not expect to find a large social multiplier in age, because people of similar age groups tend to sort together. Thus, the presence of a large number of young smokers in a particular locale would not have a large impact on the smoking habits of older people. With this in mind, we focus most heavily on our key variable—the presence of smoking bans—and look at whether the impact of this variable increases at higher levels of aggregation. We also look at the spillovers associated with years of education, income, and basic demographics (age and gender).

Table 5.5 shows the results of this estimation. The first column of table 5.5 shows our basic individual level specification. The coefficient is similar to table 5.2, though slightly larger, reflecting the restriction to 2001 and the compression of education into a single variable. The second and third columns repeat this specification at the metropolitan area level and the state level. The coefficient on the smoking ban variable increases across columns.

Table 5.5 The spillover effects of smoking

Independent Variable	Individual	MSA	State
Smoking ban	–0.061	–0.257	–0.713
	(0.007)***	(0.112)**	(0.312)**
Years of education	–0.013	–0.011	0.010
	(0.001)***	(0.009)	(0.026)
Log (income)	–0.053	–0.156	–0.271
	(0.003)***	(0.039)***	(0.082)***
N	64,660	243	51
R^2	0.05	0.26	0.59

Notes: Data are from CPS June 2001 Tobacco Supplement Survey. Sample composition is respondents eighteen years and older. Regressions weighted by self-response supplement weight. Regressions include controls for age, gender, employed, indoor worker, and a dummy for missing income. For years of education, first, second, third, and fourth grades were averaged to 2.5 years. Fifth and sixth grades were averaged to 5.5 years, seventh and eighth grades were averaged to 7.5 years, high school diploma and GEDs were treated as 12 years, some college and associates degrees were treated as 14 years, bachelors degrees were treated as 16 years, masters degrees were treated as 18 years, professional degrees (such as MD's, DD's) were treated as 20 years, and doctorate degrees (such as PhD's or EdD's) were treated as 21 years. For income, $<$ \$5,000 was coded as \$2,500, and $>$ \$75,000 was coded as \$75,000. All other categories were averaged over the range in the choice.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

The individual coefficient of –.061 increases to –.257 at the metropolitan area level and –.713 at the state level.

A social multiplier of four at the metropolitan area level and twelve at the state level gives us another estimate of 1/(1 – b), which is again compatible with an estimate of "*b*" ranging from .75 to .9. Of course, just as the variance estimates can potentially be biased by omitted area level characteristics, the social multiplier numbers are also likely to be biased upwards. Nonetheless, this provides suggestive support for significant social interactions in the smoking.

Perhaps the other two most natural candidates for variables in which to look for social multipliers are income and education. The years of education measure shows essentially no social multiplier. The logarithm of income shows a much stronger social multiplier of three at the metropolitan level and five at the state level. Again, this is compatible with high levels of social interactions, between .67 and .8.

Table 5.6 looks at these social multipliers within education categories. In this case, we just look at the social multiplier on the smoking ban variable. We find the largest social multipliers for high school graduates and the smallest for college graduates. In these regressions, social influence in smoking is more important for less educated people.

Table 5.6 **Spillover effects by education**

	Impact of smoking ban		
Education group	Individual	MSA	State
< High school	–0.028	0.075	–0.859
	(0.025)	(0.202)	(0.469)
High school grad	–0.059	–0.223	–1.303
	(0.013)***	(0.129)*	(0.423)***
Some college	–0.081	–0.426	–0.573
	(0.013)***	(0.123)***	(0.320)
College grad	–0.027	–0.075	–0.347
	(0.011)**	(0.079)	(0.187)

Notes: Data are from CPS June 2001 Tobacco Supplement Survey. The sample is individuals aged eighteen and older. Regressions are weighted and control for age, gender, employed, and indoor working.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

5.6 The Smoking Time Series

In the previous two sections, we focused on cross-sectional implications of social interactions. In this section, we turn to the dynamic implications of social interaction models and their connection with the time series of cigarette consumption. The basic structure of dynamic social interactions models is to assume that the rate at which individuals choose a behavior is an increasing function of the share of the population that is already selecting that behavior.

For example, if the population was fixed and infinitely lived, and if people who started smoking never stopped, then a dynamic social interaction model might take the form:

$$S(t+1) - S(t) = (a_0 + a_1 S(t))(1 - S(t)), \quad (2)$$

where $S(t)$ is the share of the population that smokes at time t and a_0 and a_1 are parameters. In this framework, all nonsmokers have some probability of switching to become smokers (a_0) and this probability increases with the share of the population that is already smoking. The parameter a_1 determines the power of the social interactions.

In this formulation, higher values of $S(t)$ are associated with a more S-shaped curve, and it is this S-shaped curve that is the hallmark of dynamic social interaction models. For example, figure 5.1 shows the time paths implied by three different values of a_1. In all three cases, we assume that $S(0) = .05$, and $a_0 = .02$. We show results for $a_1 = .1$, $a_1 = .2$ and $a_1 = .3$.

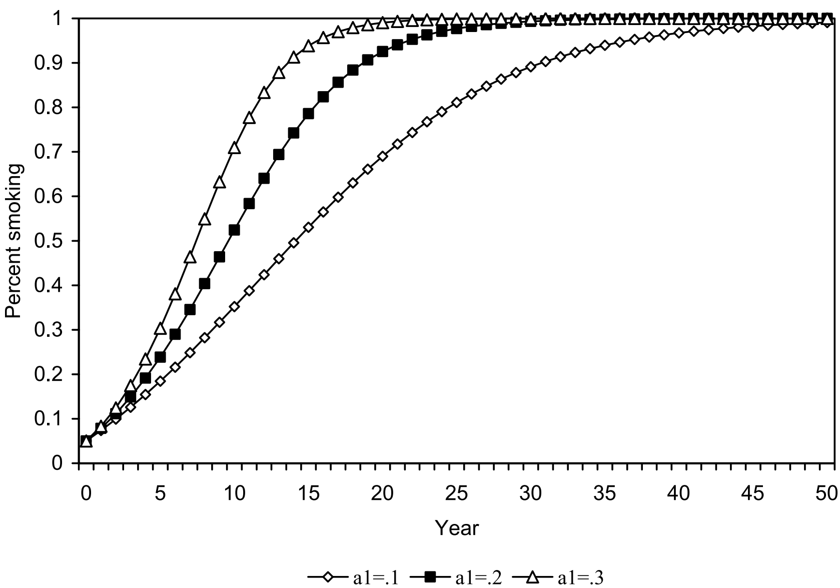

Fig. 5.1 Simulated smoking rate with initiation only

Higher values of a_1 imply both a faster convergence to everyone smoking and also a more *s*-shaped curve.

While this one-sided model might be appropriate for a time period when smoking was rising—the first half of the century, for example—it seems ill-suited for the last forty years, when cigarette smoking has been declining. A more sensible model might assume that both smokers and nonsmokers have a probability of transitioning into the other group. For example, we might assume that a nonsmoker becomes a smoker between time t and $t + 1$ with probability $a_0 + a_1S(t)$, and a smoker becomes a nonsmoker between time t and time $t + 1$ with probability $b_0 + b_1(1 - S(t))$. In this formulation, both the constant transition probabilities and the social impacts of smoking may differ. A particularly natural assumption might be that $a_1 = b_1$ so that the social impacts of smoking and nonsmoking are identical, but that the basic transition probabilities (a_0 and b_0) differ. We think of changes in beliefs about the health consequences of smoking as reflecting changes in those parameters.

With these assumptions, the new difference equation characterizing smoking rates is:

$$(3) \quad S(t + 1) - S(t) = (a_0 + a_1S(t))(1 - S(t)) - (b_0 + b_1(1 - S(t)))S(t).$$

The change in smoking includes nonsmokers who become smokers (the first term on the right-hand side) and smokers who become nonsmokers (the second term on the right-hand side). This equation can be rewritten:

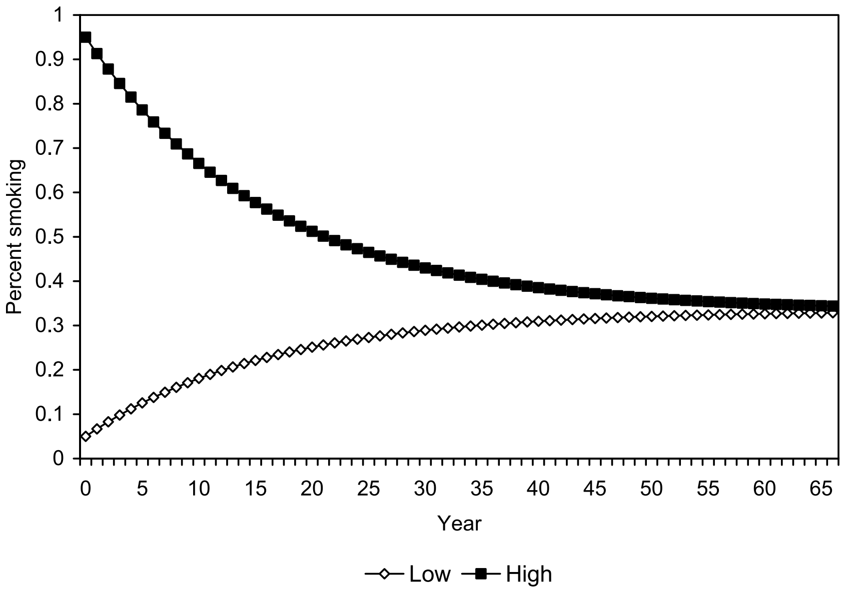

Fig. 5.2 Simulated smoking rate with quitting and initiation

$S(t + 1)–S(t) = a_0 + (a_1 – a_0 \quad b_0 – b_1)S(t) + (b_1 – a_1)\mathrm{S}(\mathrm{t})^2$. In the case that social interactions are the same for smoking and nonsmoking ($a_1 = b_1$), this equation reduces to:

$$S(t + 1) - S(t) = a_0 - (a_0 + b_0)S(t). \tag{4}$$

In this case, the system will converge to a steady-state $\overline{S} = a_0/(a_0 + b_0)$. Figure 5.2 shows two cases where the impact of social interactions on both transitions is the same.[3] In the first case, $S(t)$ converges to the steady-state level from below (starting at a smoking rate of 5 percent), and in the second case it converges from above (starting at a rate of 95 percent).

Our modest empirical implementation of this is to regress changes in the smoking rates since 1965 on the initial share of the population that smokes and the square of that share. We use data from 1965 because that is when data on adult prevalence are first available from the National Health Interview Survey (NHIS). Not all years of the NHIS asked about smoking; we use data from all the years that do, and consider adjacent years of the data. When we estimate this equation, we find:

$$\underset{(5.57)}{\Delta\text{Rate} = 2.88} - \underset{(.38)}{.23} \times \text{Lagged Rate} + \underset{(.56)}{.004} \times \text{Lagged Rate}^2. \tag{5}$$

3. We assume $a_0 = .02$ and $b_0 = .04$, so that the steady-state smoking rate is 33 percent.

Standard errors are in parentheses. There are eighteen observations and the r-squared is essentially zero (2 percent). Changes in the smoking rate over the past twenty-five years are uncorrelated with the initial level. This seems to suggest that social interactions operate weakly at an aggregate level, though clearly the number of observations makes us cautious of drawing strong conclusions.

5.7 Conclusion

This chapter discusses the possible reasons why the decision to smoke might depend on the smoking decisions of one's peers, and the empirical implications of social interactions in smoking. The most obvious implication is that exogenous forces that make one person's smoking less likely will decrease the probability that a peer will also smoke. Other implications are that social interactions will create high levels of variance across aggregates, and that there will be social multipliers, where exogenous attributes matter more at higher levels of aggregation.

We found that individuals whose spouse faced a workplace smoking ban were less likely to smoke themselves. The instrumental variables estimate of the impact of spousal smoking suggests a 40 percent reduction in the probability of being an individual smoking if a spouse quits. These impacts were greatest for people with modest levels of education, although not uniformly so. The variance in smoking rates across states and metropolitan areas is about seven times higher than it would be if there were no social interactions and if there were no exogenous variables differing across space. We also find a significant social multiplier in the impact of smoking bans. The bans have a much stronger impact at higher levels of aggregation.

These results suggest that policy interventions that impact an individual's smoking habit will have both direct effects and also indirect effects on the smoking of peers. Workplace bans seem not only to have reduced worker smoking but also the smoking of the worker's spouse. Our results also suggest that interventions are likely to have larger impacts when they are imposed at higher levels of aggregation, although we found little evidence suggesting that social interactions can explain the shape of the time series of smoking rates.

References

Brock, W. A., and S. N. Durlauf. 2001. Discrete choice with social interactions. *The Review of Economic Studies* 68 (2); 235–60.

Case, A. C., and L. F. Katz. 1991. The company you keep: The effects of family and

neighborhood on disadvantaged youths. NBER Working Paper no. 3705. Cambridge, MA: National Bureau of Economic Research, May.
Christakis, N. A., and J. H. Fowler. 2007. The spread of obesity in a large social network over 32 years. *New England Journal of Medicine* 357 (4): 370–79.
Crane, J. 1991. The epidemic theory of ghettos and neighborhood effects on dropping out and teenage childbearing. *American Journal of Sociology* 96 (5): 1226–59.
DeMarzo, P. M., D. Vayanos, and J. Zweibel. 2003. Persuasion bias, social influence, and unidimensional opinions. *Quarterly Journal of Economics* 118 (3): 909–68.
Ellison, G., and D. Fudenberg. 1993. Rules of thumb for social learning. *Journal of Political Economy* 101 (4): 612–43.
Evans, W. N., M. C. Farrelly, and E. B. Montgomery. 1999. Do workplace smoking bans reduce smoking? *American Economic Review* 89 (5): 729–47.
George, L., and J. Waldfogel. 2003. Who affects whom in daily newspaper markets? *Journal of Political Economy* 111 (4): 765–84.
———. 2006. The *New York Times* and the market for local newspapers. *American Economic Review* 96 (1): 435–47.
Glaeser, E., B. Sacerdote, and J. Scheinkman. 1996. Crime and social interactions. *Quarterly Journal of Economics* 111 (2): 507–48.
———. 2003. The social multiplier. *Journal of the European Economics Association* 1 (2): 345–53.
Kuziemko, I. 2006. Is having babies contagious? Fertility peer effects between adult siblings. Princeton University. Unpublished Manuscript.
Manski, C. 1993. Identification of endogenous social effects: The reflection problem. *Review of Economic Studies* 60 (3): 531–42.
Sacerdote, B. 2001. Peer effects with random assignment: Results for Dartmouth roommates. *Quarterly Journal of Economics* 116 (2): 681–704.
Topa, G. 2001. Social interactions, local spillovers and unemployment. *Review of Economic Studies* 68 (2): 261–95.

Comment Arie Kapteyn

It is gratifying to see that economics is catching up quickly with other social sciences (particularly sociology and social psychology) by incorporating social interactions into models of behavior. One contribution economics may make is to bring rigor to the field and to characterize in particular what is and what is not identifiable in the models that we consider (as in Manski [1993]). Smoking is clearly an example where we would expect social interactions to be important, but also one where social interactions are hard to distinguish from other reasons for observing clusters of smokers or nonsmokers. The most obvious problem is that smokers (or nonsmokers) may flock together. So if we see that smokers often have friends or spouses who smoke, this may point to social interactions, but it may also simply indicate correlated preferences.

Arie Kapteyn is a senior economist at RAND Labor and Population.