Distinguishing Spurious and Real Peer Effects:

Evidence from Artificial Societies, Small-Group Experiments,

and Real Schoolyards

Robert MacCoun*

University of California, Berkeley

Philip J. Cook, Clara Muschkin, and Jacob Vigdor

Duke University

Abstract: In a variety of important domains, there is considerable correlational evidence suggestive of what are variously referred to as social norm effects, contagion effects, information cascades, or peer effects. It is difficult to statistically identify whether such effects are causal, and there are various non-causal mechanisms that can produce such apparent norm effects. Lab experiments demonstrate that real peer effects occur, but also that apparent cascade or peer effects can be spurious. A curious feature of American local school configuration policy provides an opportunity to identify true peer influences among adolescents. Some school districts send 6th graders to middle school (e.g., 6th-8th grade "junior high"); others retain 6th graders for one additional year in K-6 elementary schools. Using administrative data on public school students in North Carolina, we have found that sixth grade students attending middle schools are much more likely to be cited for discipline problems than those attending elementary school, and the effects appear to persist at least through ninth grade. A plausible explanation is that these effects occur because sixth graders in middle schools are suddenly exposed to two cohorts of older, more delinquent peers.

1 INTRODUCTION

The vast empirical literature on deviance consistently finds large and reliable associations between an actor's behavior – be it drinking, smoking, drug taking, premarital sex, or delinquency – and that of his or her peers. In contrast, the smaller but still sizeable empirical literature on criminal deterrence finds weak and inconsistent associations between these behaviors and actual or perceived legal sanction certainty and severity.

* Corresponding author: Robert MacCoun, Goldman School of Public Policy and Boalt Hall School of Law, University of California at Berkeley, 2607 Hearst Ave., Berkeley, CA 94720-7320, <maccoun@berkeley.edu>

(These literatures are reviewed in MacCoun and Reuter, 2001.) Considered together, it is tempting to interpret these patterns as support for the notion that the behavior not under formal social control is instead under informal social control – just not in the direction we might like.

But in the 1990s a number of investigators independently concluded that much of the correlational evidence for "peer effects" was methodologically inconclusive (see Aseltine, 1995; Ennett and Baumann, 1994; Kandel, 1996; and especially Manski, 1995, 2000).[1] The problem of inferring causation from correlation will be familiar to most readers and will not be rehearsed here. But the fact that a correlation is vulnerable to rival hypotheses does not itself prove that a causal inference is invalid. There are strong theoretical and empirical reasons for believing that peer influence is genuine and important. Thus, rather than focusing on the methodological issues in identifying true peer effects, in the next two sections, we illustrate why rival processes are theoretically plausible and empirically common.

The most convincing method for establishing true peer effects is the small-group experiment. Such studies show conclusively that true peer effects occur in a variety of tasks and settings. But some experiments also illustrate that apparent peer effects can be spurious. And the procedures necessary to gain experimental control often produce situations that differ in important ways from naturalistic social settings of interest.

An alternative is to look for so-called "natural experiments" where natural or political factors relatively uncorrelated with deviance causation create exposure to peers in ways that would not otherwise occur. The quotes on "natural experiment" are an important reminder that such situations are not true experiments with true random assignment. But these situations afford stronger causal inference than would be found in a typical cross-sectional or panel survey study. As an illustration, the final section of the paper presents results from a study of the effects of administrative school configuration on adolescent behavior. In the United States, through various accidents of history, some school districts send sixth graders off to "middle schools," where they are the youngest cohorts and are suddenly exposed to older peers. In other districts, sixth graders remain in elementary schools for one more year, where they are the oldest cohorts. This accident of local history provides an opportunity to disentangle sorting and time effects from direct peer influences. We believe the analysis provides convincing evidence for very strong and important genuine peer effects, and that the results have important educational policy implications.

2 REAL AND SPURIOUS PEER EFFECTS

This paper emphasizes the term "peer effect," but the arguments about potential spuriousness probably apply with equal force to correlational evidence for any of a number of different labels, including "norm effect," "contagion," "diffusion," "bandwagon," and/or

---

[1] Interestingly, very similar methodological concerns have been raised in the literature on the identification of true "culture" in chimpanzees and other non-human organisms; e.g., Whiten, Horner, and de Waal (2005).

"information cascade" effects. There is no consensus on a single definition of a "peer effect" or "norm" in the social sciences; the terms seem to denote fuzzy sets of informal social influences that include a variety of different normative processes and mechanisms, as suggested by Table 1 (Feldman and MacCoun, 2005; also see Cook and Goss, 1996).

Table 1. Approximate overlap among social influence taxonomies.

Kelman	Compliance		Identification	Internalization*		
French & Raven	Coercive power	Reward power	Reference power	Legitimate power	Expert power	(Argument strength)**
Cialdini	Descriptive norms		Injunctive norms			
Petty & Cacioppo	Strength in numbers; strength of other peripheral persuasion cues					Strength in arguments

Note: Reproduced from Feldman, Y., and MacCoun, R. J. (2005). Some well-aged wines for the "new norms" bottles: Implications of social psychology for law and economics. In Francesco Parisi and Vernon Smith (eds.), *The law and economics of irrational behavior* (pp.358-394). University of Chicago Press. This table reflects our interpretation but is adapted from the work of Cialdini et al. (1993), French and Raven (1959), Kelman (1961), Petty and Cacioppo (1986), and Shaver (1987).

Mechanisms that most theorists would identify as "true peer effects" – in the sense of a direct causal influence of one's peers on one's behavior -- include:

- Other people as a source of approval and disapproval, or of more tangible rewards and punishments;
- Other people as source of information about "why to do X," "how many people are doing X," "how to do X," and "where and when to do X;" and
- "Stigma swamping" (Caulkins and MacCoun, 2003) in which the sheer prevalence of a behavior reduces the informal stigma associated with deviance; like Skinner's "negative reinforcement," this is the removal of a punisher rather than the presence of a reward.

But there are also a variety of mechanisms that can produce "normlike" correlations, but would not be characterized as true norm effects, include:

- Peers as "occasions" for doing X (a sex partner, a target for victimization, etc.);
- Peers as providers of enabling resources (drugs, weapons, etc.);
- "Enforcement swamping" (Kleiman, 1993) in which the sheer prevalence of a behavior overwhelms enforcement agents, reducing the risk of sanctioning;

- "Pseudomomentum effects" in which actors change independently of each other, due to common exposure to non-peer sources and/or common developmental changes; and

- Sorting or selection effects (assortative mating, differential association, "birds of a feather," labelling effects).

To say that the latter mechanisms produce "spurious peer effects" is not to suggest that they are unimportant. On the contrary, these mechanisms are of both theoretical and policy interest in their own right. But they are not direct causal influences, and misidentifying them as norm effects risks theoretical confusion and misguided suggestions for intervention.

3 SORTING EFFECTS: EVIDENCE FROM ARTIFICIAL SOCIETIES

Sorting and selection are the most widely recognized mechanisms for a spurious norm effect (Aseltine, 1995; Ennett and Baumann, 1994; Kandel, 1996; Manski, 1993, 2000; Mercken et al., 2007). The idea is simple: Correlations between self-reports of own and peer behavior may simply indicate that "birds of a feather flock together" – people tend to associate with like-minded (and like-acting) others. Although the idea is familiar, it may be helpful to provide a graphical illustration of just how difficult it can be to distinguish social sorting from social influence. We can do so by using simulated "artificial societies," in which, by design, only influence is possible or only sorting is possible, but not both.

Figure 1 depicts two such artificial societies, represented by lattices of 1500 individuals on a 51 x 51 cell lattice. The lattice on the left depicts a "Change" model, in which actors cannot move, but will change their type if enough of their neighbors are of the other type. Although the specific details differ (as explained in the Appendix), this model is conceptually similar to cellular automata models of culture developed by Axelrod (1997), Epstein and Axtell (1996), and Latane (1996). The lattice on the right depicts a "Move" model, in which actors cannot change types, but will move to another location if enough of their neighbors are of the other type. Again, while the details differ, this model is similar to Schelling's (1971, 1978) classic analysis of residential segregation, and the many variants that have been published since his original articles.

Other than the different color schemes – which we use to make it easier to distinguish the two models – the plots show a quite similar qualitative pattern. In one case, agents have "converted" to become more like the neighbors in their fixed location; this is a true norm effect. In the other case, agents have moved locations until they have found neighbors that are more similar to themselves – a sorting effect. The result in each case is a set of intertwined, dendritic social clusters. The locations in the lattices need not correspond to physical space; rather, they can be thought of as locations in a sociocultural space characterized by attitudes and behaviors. If an investigator were to collect data from either "society" at a single point in time, it would be impossible to determine whether sorting or influence had produced the observed cultural configuration.

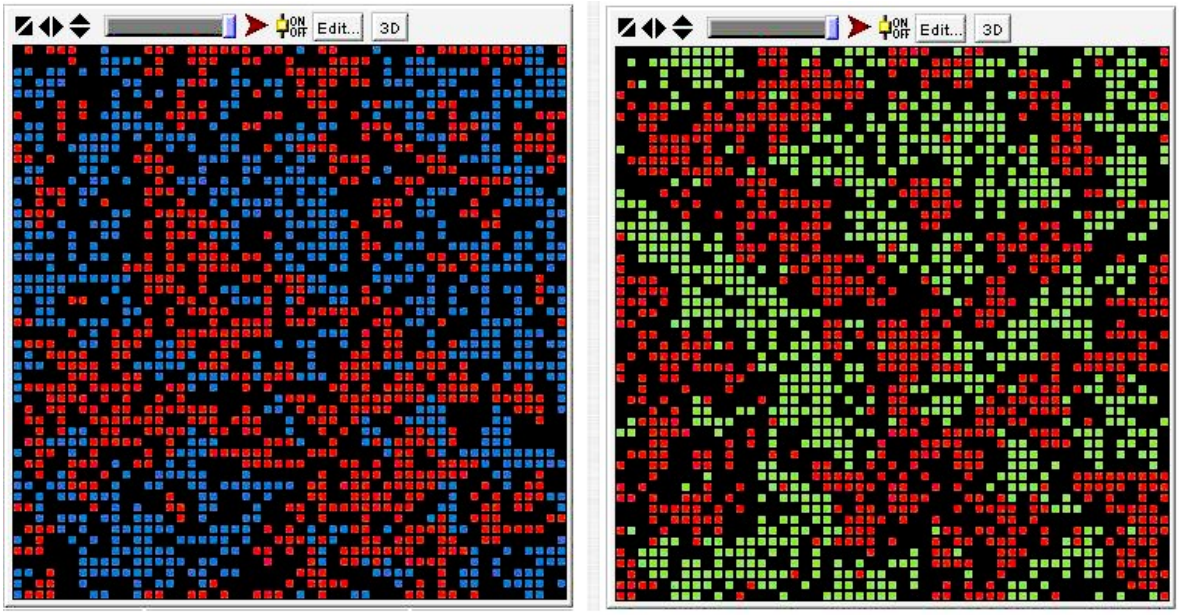

Figure 1: Social clustering after 100 iterations of two types of agents (N=1500) that were initially distributed randomly,. Left panel: "Change" model. Right panel: "Move" model.

4 PSEUDO-MOMENTUM EFFECTS: EVIDENCE FROM THE LABORATORY

Another mechanism producing "pseudonorm" effects is perhaps less obvious, but no less important. Often investigators use longitudinal rather than cross-sectional data to infer norm effects; this is particularly common in the literatures on "contagion," "bandwagon," or "cascade" effects. One commonly cited "signature" of such effects is a momentumlike pattern, in which changes in attitudes or behavior at time t are associated with an increased likelihood of similar changes at time t+1, at least among those still susceptible to change. But again, the purely correlational nature of the evidence precludes strong causal inference.

There is little doubt that such effects can and do occur in the real world, but our most rigorous and convincing evidence comes from small-group experiments, which allow researchers to experimentally control the influence process. Sherif (1948) first demonstrated contagion-like processes showing social convergence in "perceptions" of the magnitude of a visual illusion; the effect could not be attributable to common exposure to information because the perceived stimulus (the movement of a dot of light) did not in fact exist. The major innovation of Asch (1956) was to systematically and parametrically vary social influence by hiring students to pose as research participants advocating particular beliefs at variance with those of the true participants. Studies in this tradition establish that reliable and nontrivial conformity effects occur (see Bond and Smith, 1996), and field experimentation has demonstrated that they can be replicated outside the laboratory in streets and other settings (Cialdini et al., 1990; Milgram et al., 1969; Mullen, Copper, and Driskell, 1990).

But controlled experiments have also made it possible to demonstrate that some momentum-like patterns are spurious. For example, Kerr, MacCoun, Hall, and Hymes (1989) conducted three experiments in which they disentangled arguments for each of two positions from movement in the relative size of each faction (the "bandwagon"). They achieved this by mimicking the "alternate juror" role that occurs in many American trials. Alternate jurors are selected at the same time as other trial jurors, sit in the same jury box, and attend the same trial. But they only deliberate with the jury if they are needed to replace a sick or disqualified juror. In such cases, they sometimes join an ongoing deliberation "midstream."

In the Kerr et al. paradigm, participants were told that they would substitute for a seated juror in a mock jury deliberation. They then heard tape-recorded highlights of the deliberation, and were told the results of the initial and most recent ballots of votes on the jury. This paradigm makes it possible to use random assignment, to create a change in arguments while holding faction size (votes) constant, or a change in votes while holding arguments constant. The study did find that participants' votes were sensitive to current faction sizes, but there was no added impact of a faction's growing support on the alternate juror's likelihood of joining that faction. The results indicated that what appeared to be momentum or bandwagon effects in jury deliberations are sometimes an illusion, caused by differences in the rate at which group members are independently influenced by the same argument or information.

In the jury case, this may seem like "splitting hairs." Though the effect was produced by "strength in arguments" rather than "strength in numbers," arguably it is still a norm or peer influence effect, since the arguments allegedly came from group members. But in many situations, the kinds of arguments or information that produces influence will actually come from *outside* the group (e.g., through the mass media). For example, studies in the diffusion of innovation literature frequently attribute momentumlike patterns to peer influence without ruling out a temporal lag in the rate at which people respond to some purely non-social source of information.[2]

Pseudo-momentum effects can also occur if actors adopt new beliefs or behaviors at slighly different points in time, simply because of maturational changes – e.g., the rebelliousness, experimentation, and boundary testing characteristic of the onset of adolescence. In the remainder of the paper, we examine a case where this concern is especially plausible.

5 REAL PEER EFFECTS: EVIDENCE FROM THE SCHOOLYARD

Sorting and pseudo-momentum effects seem especially plausible in school settings, especially in late childhood and early adolescence. Sorting occurs formally through class assignments, but more strikingly through the fiercely enforced social cliques that are the

---

[2] E.g., in the large literature on the Bass diffusion model, correlated movement among individuals is attributed to an "imitation" coefficient, and only the residual variance is attributed to "innovation" or to external information (see MacCoun, 2007).

focus of so many Hollywood movie and television plots. Momentumlike patterns also occur because there are fairly predictable behavioral changes associated with adolescence, but they occur earlier for some students than for others. It is nearly impossible to convincingly distinguish such effects from true peer-to-peer influences using purely correlational data.

But variation in idiosyncratic institutional decisions made by local school boards create an inferential opportunity.[3] At the beginning of the twentieth century, school configuration in the United States began moving away from an eight-year primary and four-year secondary model, toward a definition of secondary education as beginning in the seventh grade (Goldin 1999). But in recent decades there has been a shift toward the middle school configuration of grades 6-8, or occasionally 5-8. In the early 1970s, less than one-quarter of middle schools incorporated sixth grade: by 2000, three-quarters of all middle schools enrolled sixth grade students.

There may be economic, logistical, and administrative reasons for configuring schools this way, but theories of normative influence suggest that spending sixth grade in a middle school might have very different behavioral consequences than would an additional year spent in an elementary school setting. Sixth graders in an elementary school will be the oldest students in the school. Sixth graders in middle school will be the youngest, with daily exposure to older adolescents. This means that they are potentially exposed to each of the normative mechanisms listed above – peer approval, peer sanctioning, peer modeling, and peer information – from an older peer group they might not otherwise encounter until a year later.

Of course, one might reasonably respond that "we can't keep them in elementary school forever," and that an administrative threshold has to be imposed somewhere. There are two counterarguments, one logical and the other psychological. Logically, the argument overlooks a basic principle of cost-benefit analysis, which is that an intervention that postpones costly behavior for a year provides a year's worth of social savings.[4] Psychologically, the argument overlooks a growing body of evidence that youth are more vulnerable to peer effects at some ages than others.

Research on adolescent delinquency suggests a developmental pattern of delinquent peer influence: the influence of peers on behavior already is significant in early adolescence, peaks during middle adolescence, and then begins to decline (Jang 1999). The transition to adolescence is a difficult time of life at best. Between the ages of 10 and 14, students typically must adjust to puberty, as well as to changes in social relationships

---

[3] The material that follows is adapted from a far more detailed presentation in Cook et al. (2007). This research is based on data from the North Carolina Education Research Data Center at Duke University, and supported by grants from the National Institute on Drug Abuse and the William T. Grant Foundation.

[4] Indeed, much of the social benefit of drug treatment programs probably comes from the reduction in drug use (and street crime) *during* treatment, even if clients quickly relapse after treatment. See Rydell, Caulkins, and Everingham (1996).

with peers, family, and authority figures (Eccles et al. et al., 1993; Elias et. al, 1985; Rudolph et al., 2001). Research suggests that difficulties in coping with multiple transitions may underlie some of the negative effects that many students experience during the transition from elementary to middle school (Eccles et al. et al., 1993). These effects include a decline in motivation and a loss of self-esteem, particularly when the transition occurs at younger ages (Rudolph et al., 2001); decline in academic achievement (Alspaugh, 2001; Hanushek, Kain and Rivkin, 2004); strains on interpersonal functioning (Barber and Olsen, 2004); and in the long term, increased risk of dropping out of school (Rumberger, 1995).

In light of these concerns, it is surprising that so little empirical research has been done on the behavioral consequences of school grade configurations. Using unique data on disciplinary infractions and end-of-grade (EOG) standardized test scores for North Carolina public school students, we exploit variation in grade configuration across and within the state's school districts. Specifically, we compare the behavioral and academic outcomes of students who attend different types of schools in sixth grade.

Our statistical analysis treats observed negative behaviors as a threshold function of an unobserved behavioral propensity. That propensity is in turn assumed to be a function of a vector of student-level characteristics and a vector of school characteristics – including the school's grade span (see Cook et al., in press). We cannot eliminate the possibility that unmeasured qualities of the students or their schools are important and distributed differently among the two groups of schools, but we attempt to minimize this risk through statistical matching and a pseudo-longitudinal analysis, described below.

Our analysis makes use of an administrative database covering all public schools and students in the state of North Carolina for a number of years. The data were provided by the North Carolina Education Research Data Center (NCERDC). The indicators of behavioral problems are derived from a statewide database of disciplinary infractions recorded during the 2000-1 academic year. Each disciplinary report reflects a decision on the part of a school official (usually a teacher) of whether to "write up" a student for misbehaving, and then a decision on the part of the principal of whether to report to the state. (Schools are required to report incidents in the event that they result in the out-of-school suspension of one or more students, or if the offense is severe enough to warrant the contact of law enforcement officials, but reporting is otherwise left to the discretion of school officials.)

In our statistical analysis we work with data from 99 school districts -- nine in which all sixth graders attend elementary school, eight in which schools differ with respect to grade configuration, and 82 where all sixth graders attend middle schools.

In our statistical work we used a matching procedure to address the concern that sixth graders are not randomly assigned to elementary or middle schools in North Carolina (see Cook et al., in press). We used logit regression to predict the likelihood that the school was a middle school on the basis of its observable geographic, economic, and demographic characteristics, and excluded schools where the imputed probability was very high (higher than for any of the elementary schools in the sample) or low. The results suggested that the

middle schools tend to have a higher concentration of blacks and Hispanics, and be located in larger, better funded districts. Rural concentration had little influence.

Our matched sample included 243 schools with 44,709 sixth graders, 11 percent of whom were in elementary schools. The sixth grade students in middle schools were less likely to be poor (as indicated by qualification for a free lunch), were more likely to have a college-educated parent, and had somewhat higher EOG scores on average. The matching procedure had the effect of reducing differences between the two groups with respect to race, per-pupil expenditures, and size of the district.

A total of almost 20,000 infractions by sixth graders were recorded in the matched sample during the school year 2000-1. Figure 2 shows that while many of the infractions were for minor events or rowdiness, violence played a prominent role. The incidence for middle school students was .47, or about 1 infraction for every two students – although in fact infractions were quite concentrated, and only 16.5 percent of students appear in the infractions database. Most notable for our purposes is that both the incidence and prevalence rate for every type of infraction were considerably higher for sixth graders in middle-school than for elementary-school students. The overall incidence was three times as high for middle school students, and the prevalence rate twice as high.

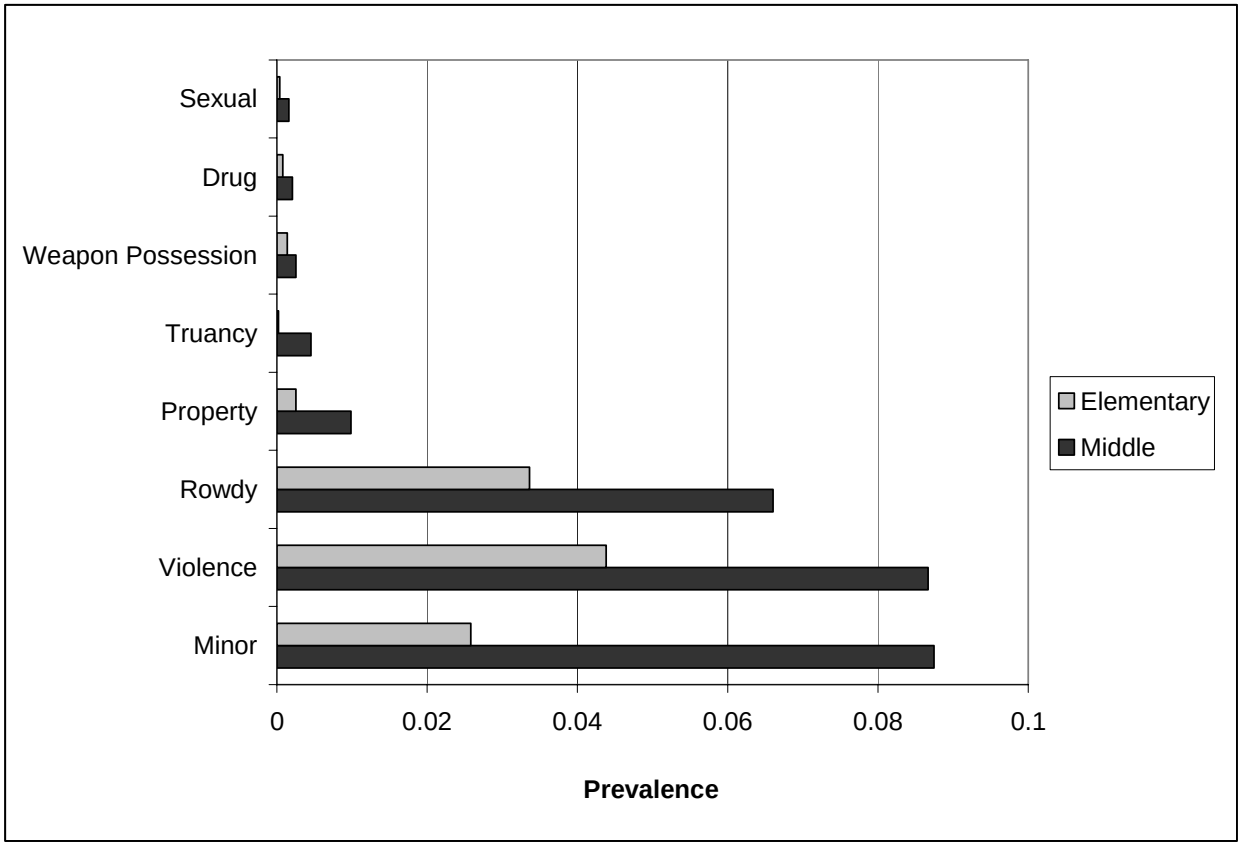

Figure 2. Infraction Prevalence - 6th Grade

The large differences in the infraction rate may actually *understate* the effect of sending sixth graders to middle school, since in our sample the middle-school sixth graders are more privileged on average. We used regression analysis in an attempt to adjust for these remaining post-match differences, and report the results in Table 2. The sample for this analysis consists of sixth grade students in North Carolina in 2000-1 that are in our matched sample. For columns 1-3, the results are from logistic regressions where the dependent variable indicates whether (1) the student appears in the infractions database; (2) the student appears in the database for a violent infraction; and (3) whether the student appears in the database for a drug-related infraction. For column 4 the results are from a negative binomial regression where the dependent variable is the number of infractions of any sort.

The results confirm that attending middle school in sixth grade is associated with greatly elevated odds of an infraction and of infraction rates. Our point estimates imply that other things equal, the odds of having at least one infraction in sixth grade are increased by a factor of 2.2 if in middle school; the odds of a violent infraction are increased by a factor of 2.1, and the odds of a drug infraction by a factor of 3.8. The results from the negative binomial regression indicate that the incidence of violations is also greatly elevated.

In these regressions, individual-level control variables include sex, race, parent's education and poverty status, old for grade, and preceding year's EOG scores. Most of these prove significant and quite influential for the prevalence and number of infractions. It is noteworthy that Hispanics tend to have lower infraction rates than non-Hispanic whites, other things equal.[5] Also included are school- and district-level characteristics, although with a few exceptions they do not prove significant. In particular it is interesting that the number of students in the sixth grade has a negligible effect on infraction rates.

We are only able to use infractions data for the single academic year (2000-1), so it is not possible to follow the behavior of individual students over time.[6] However, we are able to perform a pseudo-longitudinal analysis of behavior based on the fact that our database, while only including one year of infractions data, does include a number of years' worth of data on other aspects of each student's career. In particular we know what sort of school the students who are in fourth or fifth grade in 2000-1 are destined to spend sixth grade, and we know in what sort of school older students in that year did spend sixth grade. Using this information, we sort all students in grades 4-9 in 2000-1 into two groups, which we identify as 6Es and 6Ms. For example, a ninth grader is a "6M" if she spent her sixth grade in middle school; a fourth grader is a "6E" if he subsequently attends sixth grade in an elementary school.

---

[5] In results not shown here, we found that the effect of placing sixth grade in middle school had more or less uniform results on different demographic groups.

[6] Infractions data are available for later years, but changes in the reporting format of the data render it considerably more difficult to match these reports to student records.

Table 2
The effect of school configuration on infractions (Matched sample)

	1. Any infraction *Logit*	2. Violent infraction *Logit*	3. Drug infraction *Logit*	4. Number of infractions *Negative binomial*
In middle school	**0.799**	**0.730**	**1.330**	**0.919**
	(0.195)	(0.189)	(0.654)	(0.210)
Male	**1.122**	**1.231**	**0.955**	**1.247**
	(0.043)	(0.052)	(0.279)	(0.042)
Race (white omitted)				
Black	**0.632**	**0.658**	-0.602	**0.619**
	(0.050)	(0.070)	(0.358)	(0.057)
Hispanic	-0.431	-0.478	0.045	**-0.480**
	(0.081)	(0.103)	(0.488)	(0.092)
Asian	-1.379	-1.155		**-1.444**
	(0.245)	(0.278)		(0.252)
Other	0.147	0.289	-0.409	0.178
	(0.115)	(0.150)	(0.613)	(0.108)
Parent's education (High school grad omitted)				
High school dropout	**0.318**	**0.298**	**0.734**	**0.306**
	(0.044)	(0.051)	(0.309)	(0.041)
Trade school	**-0.199**	-0.199		-0.154
	(0.085)	(0.103)		(0.121)
Community college	**-0.143**	-0.132	-0.214	**-0.222**
	(0.059)	(0.074)	(0.431)	(0.065)
4-year college	**-0.489**	**-0.640**	**-0.923**	**-0.563**
	(0.067)	(0.081)	(0.450)	(0.067)
Graduate degree	**-0.865**	**-1.048**		**-0.923**
	(0.156)	(0.207)		(0.179)
Reduce/free lunch	**0.436**	**0.408**	0.505	**0.499**
	(0.043)	(0.050)	(0.263)	(0.044)
Old for grade	**0.372**	**0.329**	**0.535**	**0.406**
	(0.044)	(0.054)	(0.200)	(0.047)
Math EOG score, 5th grade	**-0.211**	**-0.203**	**-0.312**	**-0.259**
	(0.029)	(0.038)	(0.157)	(0.032)
Reading EOG score, 5th grade	**-0.204**	**-0.141**	-0.158	**-0.192**
	(0.028)	(0.034)	(0.143)	(0.028)
School-level variables				
% reduced/free lunch	-0.615	0.050	-1.235	-0.866
	(0.632)	(0.561)	(1.777)	(0.695)
% black	0.850	**0.831**	-1.109	**1.160**
	(0.460)	(0.407)	(2.076)	(0.528)
% Hispanic	-0.527	-2.578	-2.774	-0.254
	(1.852)	(1.584)	(5.094)	(1.653)
% parents without HS diploma	-0.300	0.467	0.453	-0.587
	(1.042)	(0.984)	(3.896)	(1.195)
District level				
Number of 6th graders	-0.053	-0.068	-0.174	-0.023
	(0.112)	(0.101)	(0.305)	(0.100)
Per-pupil expenditure, local	**0.572**	**0.616**	-0.052	0.408
	(0.237)	(0.231)	(0.602)	(0.272)

Per-pupil expenditure, federal	-0.356 0.488	**-1.066** (0.541)	0.685 (1.343)	-0.532 (0.539)
Constant	**-3.806** 0.562	**-4.589** (0.527)	**-7.670** (1.513)	**-3.214** (0.593)
Sample size	44,709	44,709	40,715	44,709

Note: bold font indicates that the coefficient estimate is significantly different from zero, p<.05. All standard errors are cluster corrected by school.

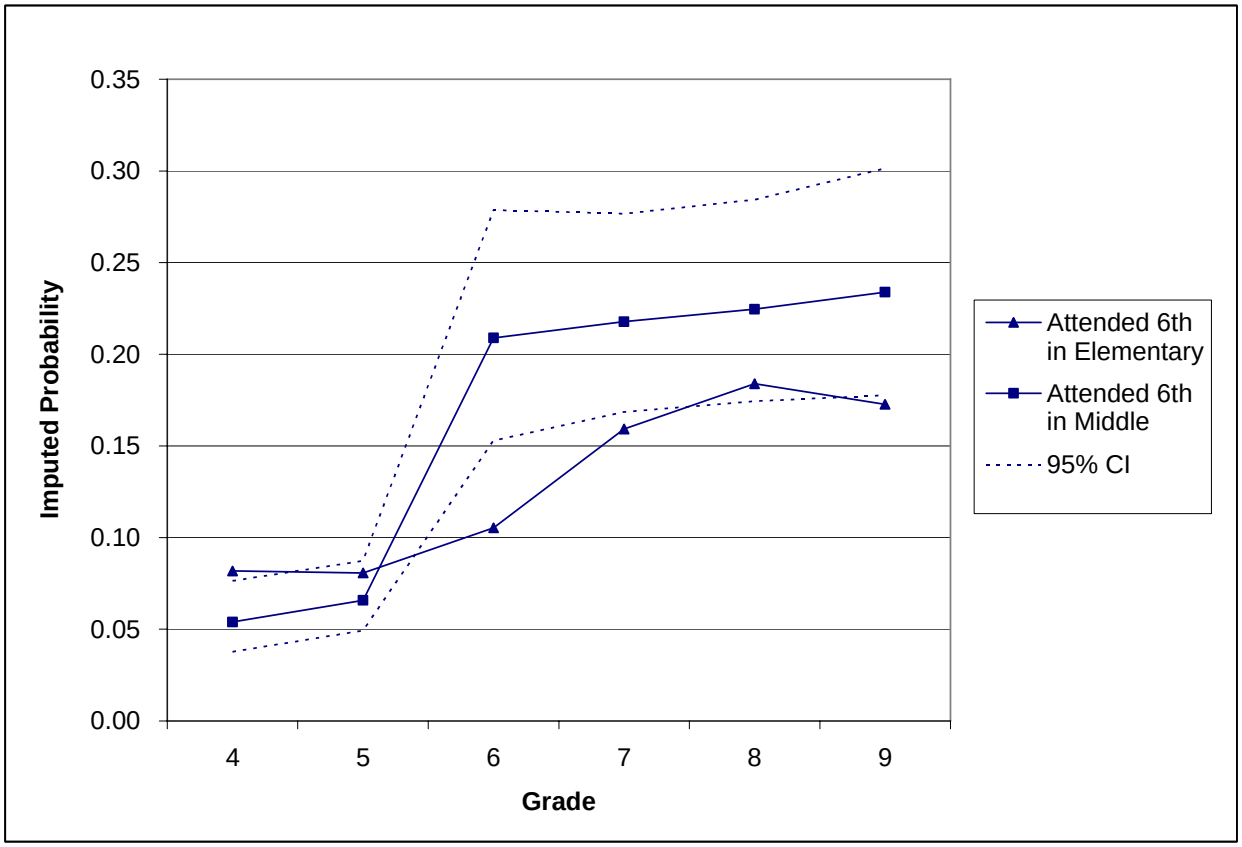

Figure 3. Probability of at Least One Infraction

Figure 3 graphs the trajectories for the two groups with respect to probability of an infraction. These prevalence trajectories are computed for the same set of values for the regression covariates, shown in Table 5; the difference in trajectories reflects the proportional effect on the infraction probability estimated from the logistic regression, and the 95 percent confidence interval represents the uncertainty in that estimate. We see that in the baseline period, grades 4 and 5, 6Es actually have a slightly higher infraction rate than 6Ms. But a large gap in the other direction opens up in sixth grade, when 6Ms have a much higher infraction rate than 6Es. The gap narrows a bit through eighth grade, at which point both 6Es and 6Ms are enrolled in middle school, a statistically significant gap persists

as far as the ninth grade. We found similar patterns in our analyses of violent and drug infractions.[7]

These results do not rule out the logical possibility that the observed differences in sixth grade are partly due to differences in school reporting practices rather than in the actual behavior of the students. It seems reasonable to suppose that middle schools tend to be more formal and severe than elementary schools, which might explain the infraction gap between 6Es and 6Ms in sixth grade. However, it does not explain why that gap persists in seventh, eighth, and ninth grades, when all the students have moved on past elementary school. Hence we believe that the observed behavior gap is not an artifact of different school reporting practices.

Further evidence for this conclusion comes from an analysis (reported in Cook et al., 2007) of statewide standardized end-of-grade (EOG) test scores for math and reading, which are not being influenced by the standards or operating procedures of the school administration. A matching and difference-in-difference procedure suggests that sixth graders in middle schools experience a drop of about 10% of a standard deviation in standardized academic achievement -- roughly equivalent in magnitude to the disadvantage associated with having an inexperienced rather than experienced teacher for a year (Clotfelter, Ladd and Vigdor 2006).

6 DISCUSSION

Decades ago the "middle school" movement was launched on the basis of plausible speculations concerning potential benefits but without much direct evidence on the effects on student behavior and performance. As it turns out, moving sixth grade out of elementary school appears to have had substantial costs. The best school configuration in which to incorporate the adolescent grades is now being reconsidered by policymakers and experts. Our results suggest that the middle school configuration that brings seventh and eighth graders into regular contact with sixth graders is problematic.

We find that despite constituting a lower-risk population along several observable dimensions, students who attend middle school in sixth grade are more than twice as likely to be disciplined relative to their counterparts in elementary school. These significant differences persist beyond the sixth grade year. Sixth graders in elementary school also make gains in standardized test scores relative to their peers in middle school. The results suggest that exposing sixth graders to older peers has negative and lasting consequences on their academic trajectories. These findings cast serious doubt on the wisdom of the historic nationwide shift to the middle school format.

---

[7] We do not trace this gap beyond 9th grade because students 16 years of age and older have the option of dropping out of school. Infraction rates decrease dramatically after 9th grade, presumably because students with the worst behavioral patterns are most likely to drop out.

The causal mechanisms that account for the inter-grade patterns of infractions and EOG scores cannot be identified directly from our data. Several differences between elementary and middle school may be relevant. In comparison with elementary school, middle school provides students more freedom and lacks the continuity and close connection provided by having one primary teacher.

But most obviously, middle school brings sixth graders into routine contact with older adolescents who are likely to be a bad influence: older adolescents as a group are more rebellious and more involved in delinquency, sex, illicit drugs, and other activities that violate school rules.

Of course the results reported here are not based on random assignment, which leaves open the possibility that the true causal process has not been adequately identified, but the consistency and strength of the findings suggests otherwise. It should also be noted that the analysis is based on data that are limited in time and place, and in particular do not include any large cities.

In response to these findings, one might ask "But don't students have to leave elementary school eventually? If sixth graders are better off in elementary school, why not seventh graders as well?" The optimal time for a transition out of elementary school will depend on many considerations, including the risks that older students pose to younger students. And of course students will differ in their developmental readiness. But it is notable that in our data, the transition was demonstrably more problematic for sixth graders than for seventh graders. There is evidence that ages 11-13 are a particularly sensitive period for the development of deviant peer influences (see Eccles et al., 1993; Haynie and Piquero 2006; Lacourse et al., 2006; Rudolf, 2001) – a time when some but not all children have entered puberty. Also, note that simply delaying adverse consequences by a year produces substantial societal benefits, assuming the transition to middle school has no effect timing of the eventual maturation out of delinquency during early adulthood.

The early sections of this paper took seriously the case that apparent peer effects might be artifactual or spurious. There are good reasons to be wary of attributing any observed cross-sectional or longitudinal peer correlations to direct peer-to-peer influence. Nevertheless, we believe our research design provides fairly compelling real-world evidence that norms scholars are quite justified in taking seriously the notion that peers can have powerful effects on each others' behavior, for better or worse. Direct normative influence appropriately deserves a prominent role in theories of social behavior, and we believe it deserves a more prominent role in consideration of public policies as well.

APPENDIX: Details of the Agent-Based Simulations

The simulation results are part of a larger modeling project described in detail elsewhere (see MacCoun, 2007), which explores the implications of a family of related nonlinear threshold models. The key model used here is:

$$p(\Delta) = (1 + \exp(-c\left[\frac{S}{N} - b\right])^{-1}$$

where p(Δ) is the probability that an agent changes views (in the "Change" model) or changes locations (in the "Move") model; *S* is the count of agents with a different belief or behavior; *N* is the total number of agents; *b* is a threshold parameter representing the critical number of opponents necessary to influence one's own belief or behavior; and *c* is a "clarity" or slope parameter which approaches a step function when *c* exceeds 10. Elsewhere, MacCoun (2007) demonstrates that this model provides a close fit to a variety of data sets in the conformity, deliberation, and diffusion literatures – generally exceeding the fit of Latane's social impact theory even after adjusting for differences in parsimony. MacCoun (2007) also demonstrates that with proper calibration, this model produces behavior nearly indistinguishable from statistical sampling models of threshold behavior (e.g., Granovetter, 1978).

Both simulations were programmed in NetLogo and used a 51 x 51 lattice with 1500 randomly placed agents -- 750 agents of each of two types (red and blue in the "Change" model; red and green in the "Move" model). The network had a torus topology (which wraps around each border) with Moore neighboods of 8 "visible" neighbors. Updating was synchronous, with 100 iterations. Simulations not presented here show that the results are reliable across runs, and only slightly influenced by topology, but that the results are profoundly effected by variations in relative levels of the threshold and clarity parameters and the proportion of neighbors that are visible to an agent (see MacCoun, 2007.)

REFERENCES

Alspaugh, J.W., and R.D. Harting. 1995. Transition effects of school grade-level organization on student achievement. 28 *Journal of Research and Development in Education* 145-49.

Asch, S. E. 1956. Studies of independence and conformity: A minority of one against a unanimous majority. *Psychological Monographs, 70* (Whole no. 416)

Aseltine, R. H. 1995. A reconsideration of parental and peer influences on adolescent deviance. 36 *Journal of Health and Social Behavior* 103-121.

Axelrod, R. 1997. *The complexity of cooperation.* Princeton, NJ: Princeton University Press.

Barber, B.K., and J.A. Olsen. 2004. Assessing the transitions to middle and high school. 19 *Journal of Adolescent Research* 3-30.

Bond, R., and P. Smith 1996. Culture and conformity: A meta-analysis of studies using Asch's (1952b, 1956) line judgment task. 119 *Psychological Bulletin* 111-137.

Caulkins, J., and R.J. MacCoun. 2003. Limited rationality and the limits of supply reduction. 33 *Journal of Drug Issues* 433-464.

Cialdini, R.B., R.R. Reno, and C.A. Kallgren. 1990. A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. 58 *Journal of Personality and Social Psychology* 1015-1026.

Cleveland, H.H., and R.P. Wiebe. 2003. The moderation of adolescent-to-peer similarity in tobacco and alcohol use by school levels of substance use . 74 *Child Development* 279-91.

Clotfelter, C.T., H.F. Ladd and J.L. Vigdor. 2006. Teacher-student matching and the assessment of teacher effectiveness. 41 *Journal of Human Resources* 778-820.

Cook, P.J. 2003. Comment on catching cheating teachers. In W.G. Gale and J.R. Pack (Eds.), *Brookings-Wharton Papers on Urban Affairs 2003* (pp. 210-215). Washington, DC: Brookings Institution Press.

Cook, P., R.J. MacCoun, C. Muschkin, and J. Vigdor. (in press). The negative impacts of starting middle school in sixth grade. *Journal of Policy Analysis and Management.*

Eccles, J.S., A. Wigfield, C. Midgley, D. Reuman, D. MacIver, and H. Feldlaufer. 1993. Negative effects of traditional middle schools on students' motivation. 93 *The Elementary School Journal* 553-74.

Elias, M.J., M. Gara and M. Ubriaco. 1985. Sources of stress and support in children's transition to middle school: An empirical analysis. *Journal of Clinical Child Psychology,* 14, 112-18.

Ennett, S. T., and K.E. Baumann. 1994. The contribution of influence and selection to adolescent peer group homogeneity: The case of adolescent cigarette smoking. 67 *Journal of Personality and Social Psychology* 653-663.

Epstein, J. M., and R. Axtell. 1996. *Growing artificial societies.* Washington, DC: Brookings Institution Press.

Feldman, Y., and R.J. MacCoun. 2005. Some well-aged wines for the "new norms" bottles: Implications of social psychology for law and economics. In Francesco Parisi

and Vernon Smith (eds.), *The law and economics of irrational behavior* (pp.358-394). University of Chicago Press.

French, J. R. P., Jr., and B. Raven. 1959. The basis of social power. In D. Cartwright (Ed.), *Studies in social power* (pp. 150-167). Ann Arbor, MI: Institute for Social Research.

Goldin, C. 1999. A brief history of education in the United States. Cambridge, MA: NBER Historical Paper No. 119.

Gootman, E. 2007a. Trying to find solutions in chaotic middle schools. *The New York Times*, Jan. 3rd.

Gootman, E. 2007b. Taking middle schoolers out of the middle. *The New York Times*, Jan. 24.

Granovetter, M. 1978. Threshold models for collective behavior. 83 *American Journal of Sociology* 1420-1443.

Hanushek, E., S. Rivkin, and J. Kain. 2004. Disruption versus Tiebout Improvement: The Costs and Benefits of Switching Schools. 88 *Journal of Public Economics* 1721-1746.

Hardy, C., W.M. Bukowski, and L.K. Sippola. 2002. Stability and change in peer relationships during the transition to middle level school. 22 *The Journal of Early Adolescence* 117-42.

Jacob, B.A. and S.D. Levitt. 2003. Catching cheating teachers: the results of an unusual experiment in implementing theory. In W.G. Gale and J.R. Pack (Eds.), *Brookings-Wharton Papers on Urban Affairs 2003* (pp. 185-220). Wash., DC: Brookings Institution Press.

Jang, S.J. 1999. Age-varying effects of family, school and peers on delinquency: a multilevel modeling test of interactional theory. 37 *Criminology* 643-85.

Kandel, D. B. 1996. The parental and peer contexts of adolescent deviance: An algebra of interpersonal influences. *26 Journal of Drug Issues* 289-315.

Kelman, H. C. 1961. Processes of attitude change. 25 *Public Opinion Quarterly* 57-78.

Kerr, N. L., R.J. MacCoun, C.H. Hansen, and J.A. Hymes. 1987. Gaining and losing social support: Momentum in decision making groups. 23 *Journal of Experimental Social Psychology* 119-145.

Kleiman,M.A.R. 1993. Enforcement swamping: A positive-feedback mechanism in rates of illicit activity. 17 *Mathematical and Computer Modeling* 65-75.

Lacourse, E., D.S. Nagin, F. Vitaro, F., S. Cote, L. Arseneault, and R. Tremblay. 2006. Prediction of early-onset deviant peer group affiliation. 63 *Arch. Gen. Psychiatry* 562-568.

Latané, B. 1996. Strength from weakness: The fate of opinion minorities in spatially distributed groups. In E. Witte and J.H. Davis (Eds.), *Understanding group behavior (Vol. 1): Consensual action by small groups* (pp. 193-219). Hillsdale, NJ: Erlbaum.

MacCoun, R.J. 2007. *Burdens of proof in social life: A formal model for integrating several major social influence paradigms.* Presented at the IGERT seminar, University of California at Berkeley, 11 April 2007.

MacCoun, R.J., and P. Reuter. 2001. *Drug war heresies: Learning from other vices, times, and places.* Cambridge University Press.

Manski, C.F. 1995. *Identification problems in the social sciences.* Harvard.

Manski, C.F. 2000. Economic analysis of social interactions. 14 *Journal of Economic Perspectives* 115-136.

Mercken, L., M. Candel, P. Willems, and H. de Vries, H. 2007. Disentangling social selection and social influence effects on adolescent smoking: The importance of reciprocity in friendships. 102 *Addiction* 1483-1492.

Milgram, S., L. Bickman, and L. Berkowitz. 1969. Note on the drawing power of crowds of different size. 13 *Journal of Personality and Social Psychology* 79-82.

Mills, R. 1998. Grouping students for instruction in middle schools. *Eric Digest*, Report: EDO-PS-98-4.

Mullen, B., C. Copper, and J.E. Driskell. 1990. Jaywalking as a function of model behavior. 16 *Personality and Social Psychology Bulletin* 320-330.

National Middle School Association (2001). Grade configuration. NMSA Research Summary #1 (updated Sept. 2001).

North Carolina Department of Public Instruction (2004). Last best chance 2004: Educating young adolescents in the twenty-first century. *Middle Grades Task Force Report*. North Carolina Department of Public Instruction and State Board of Education.

Petty, R.E., and J.T. Cacioppo. 1986. The elaboration likelihood model of persuasion. 19 *Advances in Experimental Social Psychology* 123-205.

Rudolph, K.D., S.F. Lambert, A.G. Clark, and K.D. Kurlakowsky. 2001. Negotiating the transition to middle school: The role of self-regulatory processes. 72 *Child Development* 929-46.

Rumberger, R.W. 1995. Dropping out of middle school: A multilevel analysis of students and schools. 32 *American Educational Research Journal* 583-625.

Rydell, C.P., J.P. Caulkins, and S.S. Everingham. 1996. Enforcement or treatment? Modeling the relative efficacy of alternatives for controlling cocaine. 44 *Operations Research* 687-695.

Schelling, T.C. 1971. Dynamic models of segregation. 1 *Journal of Mathematical Sociology* 143-186.

Schelling, T.C. 1978. *Micromotives and macrobehavior.* W. W. Norton and Co.

Shaver, K.G. 1987. *Principles of social psychology*, 3rd ed. Lawrence Erlbaum Associates, Hillsdale, NJ.

Sherif, M. 1948. *An outline of social psychology*, New York: Harper.

Whiten, A., V. Horner, and F.B.M. de Waal. 2005. Conformity to cultural norms of tool use in chimpanzees. 437 *Nature* 737-740.

Zernicke, K. 2007. The preteen: Betwixt and bedeviled. *The New York Times*, Jan. 7.