Developmental Psychology
2008, Vol. 44, No. 2, 355–364

0012-1649/08/$12.00 DOI: 10.1037/0012-1649.44.2.355

Using Technology to Explore Social Networks and Mechanisms Underlying Peer Effects in Classrooms

Jonathan Guryan
University of Chicago and National Bureau of Economic Research

Brian Jacob
University of Michigan and National Bureau of Economic Research

Eric Klopfer and Jennifer Groff
Massachusetts Institute of Technology

Peer interactions among children have long interested social scientists. Identifying causal peer effects is difficult, and a number of studies have used random assignment to produce evidence that peers affect each other's outcomes. This focus by sociologists and economists on *whether* peers affect each other has not been matched by direct evidence on *how* these effects operate. The authors argue that one reason for the small number of studies in sociology and economics on the mechanisms underlying peer effects is the difficulty of collecting data on microinteractions. They argue technology reduces data collection costs relative to direct observation and allows for realistic school activities with randomly assigned peers. The authors describe a novel strategy for collecting data on peer interactions and discuss how this approach might shed light on mechanisms underlying peer influence. The centerpiece of this strategy is the use of handheld computers by middle and high school students as part of interactive math and science lessons called the Discussion Game. The handhelds collect data on interactions between students and track how students' answers evolve as they interact with different peers.

Keywords: technology, peer effects, social networks, education

Understanding peer interactions and their effects on child development has interested social scientists for a long time. How does one child influence another's thoughts or behaviors? How do peer influences vary across contexts? What do the microinteractions that ultimately generate this influence and alter thoughts and behaviors look like? These issues are not merely of academic interest. The nature and extent of peer influences have important implications for policies ranging from school choice to housing assistance. In *The Truly Disadvantaged*, William Julius Wilson (1987) argued that concentrated poverty has an especially damaging impact on individual outcomes. Although the correlation between neighborhood poverty and negative outcomes is compelling on its face, these associations confound important determinants of residential location that likely also affect economic and developmental outcomes.

The difficulty associated with identifying peer effects is well-recognized (see, e.g., Manski, 1993; Moffitt, 2001), and a number of important studies have used random assignment of peers to produce good evidence that peers affect each other's outcomes (e.g., Duncan, Boisjoly, Kremer, Levy, & Eccles, 2005; Kling, Liebman, & Katz, 2007; Sacerdote, 2001; Zimmerman, 2003). This focus on whether peers affect each other has not been matched by direct empirical investigations by sociologists or economists on how these peer effects operate. Most peer effect studies by sociologists and economists treat roommates, classmates, or neighbors as peers but do not investigate the many individual interactions that define what it means to be a "peer." We argue that one reason for the small number of studies by sociologists and economists on the mechanisms underlying peer effects is that it is very difficult to collect data on these microinteractions.

The developmental psychology literature takes a step further by noting multiple ways of identifying "friends" and "peers" and by recognizing that these relationships can vary at a qualitative level in terms of content, constructiveness, and closeness (Hartup, 1996). Developmental psychologists have also focused much more than economists or sociologists on the microinteractions among children and how these affect development. Many developmental-

Jonathan Guryan, Graduate School of Business, University of Chicago and National Bureau of Economic Research, Cambridge, Massachusetts; Brian Jacob, Gerald R. Ford School of Public Policy, University of Michigan and National Bureau of Economic Research; Eric Klopfer, Urban Studies and Planning and Teacher Education Program, Massachusetts Institute of Technology; Jennifer Groff, Teacher Education Program, Massachusetts Institute of Technology.

We thank an anonymous donor for generous funding to support this research and the National Science Foundation (SES-0312749) for funding to support this research. Members of the Young Faculty Leaders Forum, including Tom Healey and Richard Light, provided important encouragement and support. We thank the individuals who helped make this work possible: the teachers of Newton Public Schools, who opened their classrooms to us and invited our out-of-the-box instructional strategies into their teaching practice; Matthew Gentzkow, Austan Goolsbee, and Susan Yoon, who made important suggestions at the early stages of this project; Judy Perry, who oversaw the design of the Discussion Game, which was implemented by students in the Massachusetts Institute of Technology Teacher Education Program.

Correspondence concerning this article should be addressed to Jonathan Guryan, University of Chicago Graduate School of Business, 5807 South Woodlawn Avenue, Chicago, IL 60637. E-mail: jonathan.guryan@chicagoGSB.edu

ists have worked to collect detailed data on interactions between students. Much of this work has been done by direct observation or by observation of videotaped interaction (see, e.g., Berndt, Perry, & Miller, 1988; Hanish, Martin, Fabes, Leonard, & Herzog, 2005; Vaughn, Vollenweider, Bost, Azria-Evans, & Snider, 2003).

In this article, we argue that new technologies can reduce the cost of such data collection. Automated data collection brings with it some obvious drawbacks when compared with direct human observation. We argue, however, that many of these may be offset by some important features of the technology.

First, because all data collection decisions must be made prior to the study, automation ensures that data collection is not affected by subconscious biases attributable to the observer. Systems of direct observations are designed to address concerns about such biases, and though they may do so reasonably well, it is reasonable to believe that automation does better on this count. Second, because the researcher does not have to be present during the data collection, the scope for experimenter demand effects (Levitt & List, 2007) is reduced. Third, the need for random assignment of peers is not obviated by the collection of detailed data on interactions. For researchers making causal inferences about peer influences, random assignment is still important (though not sufficient). We argue that data-collection technology can be integrated into the classroom setting so that peers can be randomly assigned while the students' perception that they are engaging in a normal classroom activity is preserved. To the extent that these technologies make implementation of random assignment easier, we should view them as a useful tool for the estimation of peer effects.

In this article, we describe a novel strategy for collecting data on peer interactions and discuss how this approach could be used to learn more about the mechanisms underlying peer influence. The centerpiece of this research strategy is the use of handheld computers (i.e., Palm Pilots) by middle and high school students as part of interactive math and science lessons. The students use the Palm Pilots (Model Z22) to engage in an activity called the Discussion Game, in which the primary action is students discussing their opinion or answer to a given statement. The handheld computers collect data on interactions between individual students and track how students' answers to questions evolve as they interact with different peers. We believe that as technology develops, it can be used to collect even more detailed data on interactions between students—helping us to answer a host of interesting questions about how students affect each other's academic and social development.

Our work was composed and conducted by an interdisciplinary team of researchers from the fields of economics, education, technology, and public policy. Thus, we acknowledge the challenges of working in an interdisciplinary space. However, we seek to leverage the power of interdisciplinary work, which can produce innovative methods, theories, and insights that may not have otherwise surfaced. The limitations to our methodology and preliminary practices are noted in the pages that follow. The purpose of this article is to describe our strategy so that other researchers and practitioners adapt and extend the work to further enhance the ability of technology to help us understand social interactions.

Selection and the Identification Problem

To motivate the research design described below, we illustrate here two important aspects of the research on peer effects and peer interactions: the *identification problem* and *underlying mechanisms*. The identification problem refers to the difficultly with identifying causal effects of one person's characteristics or actions on another's behavior, outcomes, or development (see, e.g., Kandel, 1978; Manski, 1993; Moffitt, 2001). For example, people choose to live in particular neighborhoods and bring their children with them. These decisions by large numbers of individuals collectively determine who a child's peers are in his or her neighborhood and school. Correlations of outcomes or developmental levels between children who live in the same neighborhood therefore confound the effects of peer interactions with the effects of the unobserved variables that factored into parents' decisions about where to live. Comparisons of outcomes of students in the same class further confound the effect of variables used by school administrators to assign students to classrooms. At a more detailed level, comparisons of outcomes between friends within a classroom are similarly confounded by the choices that led the children to become friends.

An increasingly common solution to the basic identification problem in peer effects studies is to rely on episodes of random assignment of peers.[1] Two similar examples are studies by Sacerdote (2001) and Zimmerman (2003). Both studies compare the academic and social outcomes of college students randomly assigned to freshman-year roommates with different characteristics. Because students do not choose their roommate, the researcher can rule out that correlations of outcomes are driven by the correlations in characteristics that lead students to become peers. One recent example in the child development literature is Duncan et al. (2005), who found that males who reported binge drinking in high school were more likely to engage in binge drinking in college if they were randomly assigned to a roommate who also binge drank in high school, but did not find similar relationships among females or in either gender for marijuana use or sexual behavior. Another example is the study of "Moving to Opportunity" by Kling et al. (2007), who found that the female children of families that were randomly offered housing vouchers experienced positive mental health and educational outcomes from moving to a lower poverty neighborhood; male children of families receiving vouchers experienced negative effects from moving to lower poverty neighborhoods.

In addition to the basic identification problem caused by selective matching of peers, there are at least two other complications with interpreting correlations of peer outcomes (see, e.g., Manski, 1993, for a formal discussion). First, it is hard to distinguish the effect of Child A's outcome on Child B's outcome from the effect of Child B's outcome on Child A's outcome. Both children might influence each other, and in turn the effect one child has on the other may affect the first child, and so on. It is therefore important to be careful making statements about causality. Manski (1993) calls this the "reflection problem." A second problem is that in the absence of peer effects, even randomly assigned peers might have

[1] It is important to distinguish studies with random assignment of peers from randomized controlled studies in which treatment children are taught skills to foster positive interactions with peers (see, e.g., Ladd & Mize, 1983). These studies address different questions and ideas, such as the link between prosocial behavior and peer acceptance, between aggression and social withdrawal and peer rejection, and the formation of friendships.

correlated outcomes if they are affected by common outside influences. Manski refers to this form of the identification problem as exposure to "common shocks." For example, students in the same class might have similarly low test scores if a dog had been barking loudly outside the classroom during the test.[2] The correlation in outcomes would be due neither to the assignment mechanism of students to classes nor to peer effects.

Mechanisms

The difficulty of the identification problem has led economists and sociologists to focus more in empirical studies of peer effects on *whether* peers affect each other rather than on *how* this influence occurs. If one takes as given that students in the same class affect each other's learning, what does one make of this fact? Without an understanding of how peers affect each other, it is hard to make policy recommendations on the basis of even well-identified estimates of reduced-form peer effects.

Most empirical studies of peer interactions in economics and sociology define peers on the basis of fairly crude variables. Students in the same classroom, residents in the same dorm room, and children who grow up in the same neighborhood are taken to be peers. However, as it is rare to have data on individual interactions between these supposed peers, it is nearly impossible to describe what it means for two children to be peers. It is probably true that a student interacts more with the students in his or her own classroom than with students in a different classroom, but this hides a great deal of important heterogeneity. Within classrooms, there are student pairs who interact more than others. These interactions take different forms, from social to intellectual conversations, from nonverbal communication to choices of the types of clothes to wear. If students affect each other's learning, it may be the complex intermingling of each of these interactions that determines exactly how. The typical peer effects study estimates the net effect of all of them.

There is a virtually unlimited set of potential theories to explain why children affect each other's development; even narrowing the focus to the school setting, the possibilities are vast (see, e.g., Johnson & Johnson, 1985; Knight & Bohlmeyer, 1990; Slavin, 1983). Developmental psychologists have posited and examined a number of potential mechanisms underlying the developmental effects of peers. These include but are not limited to theories of peer rejection (Coie, Lochman, Terry, & Hyman, 1992; DeRosier, Kupersmidt, & Patterson, 1994; Hymel, Rubin, Rowden, & LeMare, 1990), peer victimization (Boulton & Underwood, 1992; Kochenderfer & Ladd 1996), friendship and peer acceptance (Ladd, Kochenderfer, & Coleman, 1996; Parker & Asher, 1993; Vandell & Hembree, 1994), social information processing (Dodge, 1986), and social learning (Ladd & Mize, 1983).[3]

As they are likely to be comparatively less familiar to the readers of this journal, we describe in more depth theories discussed in the economics and education literatures. One set of theories focuses on the scarcity of the teacher's attention. Lazear (2001) argued that a behavioral outburst by an individual student forces the teacher to divert his or her attention away from pedagogy and therefore away from the rest of the students in class. Peers of students with behavioral problems therefore suffer, even in the absence of any direct interactions between students. Another possible mechanism by which peers could affect each other's learning operates through the choice of curriculum and teaching style. The learning styles and aptitudes of every student in the class may affect the teacher's choice of pedagogical technique and content. The way material is taught is therefore affected by the characteristics of other students in the classroom. As a result, each student's learning may be affected by his or her peers, though indirectly and in possibly complicated ways. The long-standing debate in education policy about ability tracking is based in part on this idea (see, e.g., Oakes, 2005). Another possibility is that peers might create a culture in which education is valued or devalued. One example of this is the debate surrounding "acting White," a phenomenon whereby Black students disengage from school in order to avoid behaving in ways that are associated with White culture (see, e.g., Cook & Ludwig, 1997; Fordham & Ogbu, 1986; Fryer & Torelli, 2005).

Hoxby and Weingarth (2006) conducted one recent empirical study that attempted to distinguish some of these mechanisms. In that study, the authors tracked how student test scores respond to having sets of peers with various combinations of characteristics. The study examined variation in peers induced by large numbers of quasirandom reassignments of public school students in Wake County, North Carolina. The number and variety of miniexperiments allowed the authors to separately identify, for example, the effects of having on average high-ability peers from the effects of having one very-high-ability peer. The authors concluded that the results support models in which homogeneity of ability within the classroom allows the teacher to target instruction toward a particular type of learner.

Another set of theories focuses directly on peer interactions. As opposed to a teacher delivering information on a one-way street to students, one might think of every student as being both a teacher and a learner. Put that way, it is clear that the ways in which students affect each other's learning must depend on who communicates with whom and how successfully. Close friends may speak to each other more and may ask each other questions about the current topic of study. Alternatively, everyone in the class may turn to one respected student to ask for help with difficult questions or to help explain confusing concepts. The important interactions may not be limited to academic ones. Students may choose to invest effort in school depending on the choices of particular peers they respect. Additionally, the development of respect itself is likely a result of many interactions, both academic and social.

As this discussion makes clear, the set of theories is too large to test in one study. Furthermore, without more detailed data on individual interactions between students than exist in typical datasets, it is impossible to test any hypotheses that these theories might generate. It is rare for researchers to collect detailed data on the interactions between students in the classroom (Gillies, 2003). We believe that technology, if properly integrated into classroom

[2] We thank Patrick McEwan for this particular example.

[3] There is also a large literature exploring the reasons why some children have more positive interactions with peers than others (see, e.g. Ladd, 1999, for a review of this literature).

activities, can help to generate data on some of these interactions.[4] In the next section, we describe a particular technology and a related activity that collects data on individual interactions between students. At this early stage of development, the technology barely begins to collect data on every important peer interaction within classrooms. However, we believe what we describe below is a good start and a possible improvement relative to moving the status quo in an important new direction.[5]

It should be stressed, however, that the collection of detailed data on interactions among peers is not a substitute for random assignment. The more data we have as researchers, the more detailed theories we can test. But we cannot make causal inferences about peer interactions without a valid research design, regardless of the available data.

The method we describe below is intended to test a subset of potential mechanisms related to information transfer between students and peer influences. Do students communicate important information related to questions they are trying to answer in class? Do students learn directly from their peers? Which types of students are most influential? Do the answers to these questions vary depending on the type of question the students are answering? In attempting to shed light on these questions, we intend to push social scientists toward answering the question of *how* peers affect each other's outcomes in the classroom, as distinct from *whether* they do. Of course, each of the questions we address masks another more detailed underlying mechanism that would require even more detailed data. It is important here to distinguish the limitations of the technology as it currently exists and is currently being used from the inherent limitations of similar technologies. Technological progress, along with improvements in the imaginations of researchers who use them, will allow for even more detailed data collection than what we describe. To start, however, we describe what currently exists and leave the future possibilities to the reader's imagination.

The Discussion Game

We now turn to a discussion of a particular use of technology in classrooms that has three broad aims: (a) to collect data on individual interactions between students in a classroom setting, (b) to explore a specific set of mechanisms by which peers might affect each others' development, and (c) to experimentally manipulate interactions between students to help identify causal effects.

We begin with a description of a particular implementation of this method. However, it should be clear that many of the specific details could be changed to facilitate the testing of different hypotheses and the collection of different data. After explaining the method we have already put into action, we turn to specific ways we envision supplementing and altering the technology, including the rules of the simulation to expand the set of questions that might be addressed.

In the normal course of the school year in a sixth-grade science classroom, students engaged periodically in an activity which we call the Discussion Game. This activity is based on the work of Susan Yoon (2008), and the associated software was developed in the Massachusetts Institute of Technology (MIT) Teacher Education Program (Klopfer & Yoon, 2005; Klopfer, Yoon, & Rivas, 2004). The Discussion Game software is one of a number of *participatory simulations* designed in the MIT Teacher Education Program. Participatory simulations use technology to augment human interactions in a specific scenario, providing more data to the user and allowing these scenarios to become more complex. Many participatory simulations add a small amount of technology to role-playing scenarios. In one example called the Virus Game, one student begins the scenario infected with a disease. As students interact using the handheld's infrared capabilities, some become infected, and others unknowingly become carriers. The students then use the data collected by the handhelds and their own observations to infer what they can about the disease, how it spreads, and what its effects are.

The Discussion Game is the one participatory simulation that does not have a role-playing component. Unlike the Virus Game, the Discussion Game displays real data of student responses in the class. The Discussion Game software was designed to be played in two forms, each with a slightly different pedagogical goal. In one form, which we call the "opinion" domain, the class is presented with a statement that is designed to elicit opinions, either positive or negative. The statement is chosen to fit with the current topic on which the class is focusing, and the activity is designed to engage students in critical debate related to the relevant curriculum. For example, if the class is finishing a set of lessons on the scientific method, the statement might be "It is OK for scientists to use animals to test the effectiveness and potential side-effects of new human drugs."

At the beginning of the activity, each student is given a handheld computer that has the Discussion Game software running. The handheld computer displays the statement that the teacher has created and programmed into the game to be the current subject of discussion. After the students enter their names, the software asks each student to enter her or his personal opinion on the statement on a scale from −5 (*strongly disagree*) to 5 (*strongly agree*), with 0 being neutral (a response the software does not permit the student to enter). Students are then asked to form pairs and are allowed to discuss the statement and the reasoning behind their personal opinions. At the beginning of the paired meeting, students exchange information by "beaming" their response to their partners, using the infrared transmission capabilities of the handheld computer. Consequently, each student's handheld computer then displays her or his own response as well as the partner's response. Students then begin to discuss the statement, explaining their rationale to their partner and offering supporting arguments. After a prescribed number of minutes, each student is again asked to enter an opinion into the handheld computer. This opinion may or may not have changed after the discussion with a classmate. To

[4] Detailed data on individual student actions and interactions in the classroom do exist. For example, McFarland (2001) observed 36 classrooms for 2 days at a time and recorded episodes of conflict and defiance between students and teachers. In other cases, researchers have video recorded classes and then coded actions and interactions while watching the video (e.g., Gillies, 2003). One benefit of the handheld technology we describe below is that it collects prescribed data automatically in a way that is not subject to ex post facto human interpretation. All decisions about data collection are made prior to the period under observation.

[5] Our use of technology to study peer interactions in schools is closely related to the use of video technology by developmental psychologists for similar purposes. One example of such a study is Berndt, Perry, and Miller (1988).

conclude the meeting, students must again exchange data on their opinions by "beaming" to one another's handheld computer. Thus the handheld computers now hold information on who interacted, how long they interacted, and their responses both pre- and post-meeting. The teacher then asks the students to pair with a different person and repeat the process, emphasizing the "beaming" or sharing of opinion data both before the discussion begins and once it has concluded.

In the other form of the Discussion Game, the opinion-eliciting statement is replaced with a question that asks students for a specific answer that, unlike an opinion statement, would seemingly have a "right" answer. Instead of entering an opinion, students must provide a numerical answer and rate their confidence in that answer. We call this the "fact" domain. The questions—based on the style of questions originally used by the famous physicist Enrico Fermi to teach his undergraduate classes—may or may not have specific numeric answers but are designed to encourage students to logically narrow down the range of reasonable answers. An example might be "How many ping-pong balls fit in the cabin of a Boeing 747 airplane?" Although there likely is a definite answer to this question, it is not one that could be determined without considerable time and effort. The purpose of using these questions with students is to get them to focus less on obtaining the precise answer than on the process of getting a good estimate and therefore being the closest one in the class to the "right" answer. In regard to our example, a student might first estimate how many ping-pong balls would fit in one seat and then estimate how many seats there are on the airplane. Additional considerations would help to generate a more accurate answer. How many balls might fit in the overhead compartment? How many overhead compartments are there in the cabin? How many balls might fit in the aisle? The question is meant to teach students to think logically through a question, much like a scientist might. Consequently, sharing the manner in which an answer is derived with a classmate can generate rich discussion about the steps one used to arrive at their answer, what seems a plausible way to estimate some of these variables, and so on.

The Fermi questions also serve as an interesting comparison to the opinion statements. They are both similar enough to be comparable and different in ways that might yield interesting comparisons.

First, how are the opinion and fact domains similar? Most obviously, in both forms of the game the student's input is numerical. Less apparently, in each case the statement or question is designed so that students might value input from every student in the classroom, though in practice they might choose to respond to some subset of them. It is not hard to imagine why a student would care to know every other student's opinion on a controversial topic, if for instance she or he wished to state an opinion designed to seem "typical" or even "extreme." However, fact-based questions can by their form create an incentive for students to value the advice of a small subset of the classroom. If instead of the Fermi questions the students were given a purely factual question, the optimal strategy might be to ask the "smartest kid" in class what the answer was and ignore everyone else's input altogether. The Fermi questions are specifically designed so that no one is likely to know the exact answer. This lack of a "right" answer leaves scope for every peer's input to be valuable, though possibly to differing degrees. It may turn out empirically that seeking out the smartest kid in the class is precisely the strategy students use to try to answer the Fermi questions, but it is unlikely that they think any one student knows the answer. The form of the question thus leaves open the possibility that students value each of their peer's input, though to differing degrees, in a way that is similar to the opinion domain.

In what important ways are the opinion and fact domains different? First, students are likely to have overlapping but different sets of motivations for choosing an opinion versus an estimate of a factual answer. A student might care what others infer about him or her as a person on the basis of an opinion on a controversial topic, but it is less likely that student will be concerned what others infer on the basis of a personal estimate. Relatedly, it is possible that distinct sets of students are viewed by their peers as being more likely to have the "right" opinion versus the "right" answer to the Fermi question. The students who are academically talented or good at problem solving may or may not be the ones who are influential in setting dominant opinions.

Furthermore, though they are designed to be different from typical questions students see on tests, for the reasons previously stated, the Fermi questions more closely resemble normal schoolwork than the opinion-eliciting statements do. As we discuss later, the data created by the Discussion Game allow us to track how students' answers change after meeting and talking with peers. Students might move their opinion or estimate closer to or farther from their partner's after interacting. It seems that this choice is likely motivated by different considerations when the students are discussing an opinion and a Fermi question. Sorting through those different motivations will be important for interpreting the results generated by this research.

The first important thing to point out about the value of the handheld technology is the type of data it can collect. The handhelds record each interaction, its time, the names of the individuals, and each student's opinion. Relative to the data available in typical datasets, this is a significant step forward. In most studies of peer interactions by economists or sociologists, an interaction is defined at the neighborhood, classroom, or dorm room level. This characterization crudely summarizes a multitude of individual interactions and is done chiefly because of a lack of data. The lack of data on individual interactions forces the researcher to rely on introspection to argue that students in the same classroom have more interactions than those in different classrooms and to ignore the types of interactions altogether. The latter weakness is what makes it particularly hard to test specific mechanisms that might explain why peers affect each other's outcomes.

The handheld technology along with the ability to manipulate the form of the classroom activity allows us to collect data on individual interactions. These are the data that we hope will allow testing of some subset of the possible mechanisms by which peer effects operate. It should be noted, however, that the technology at this stage is limited. We do not observe every detail of every interaction between students. For example, we do not record what each student says during a conversation. Nor do we observe what students say to those with whom they are not officially paired. We hope to develop technology that allows students to enter particular arguments they made to support their points, but even this would be an imperfect record of the conversation. Furthermore, body language might be important, as might be every previous interaction that the two students have ever had. Clearly, technology can only hope to capture so much of the important information transferred between the peers. We consider the Discussion Game a first step toward collecting the kind of detailed data that can help to answer interesting questions about peer interactions.

One important feature that we mentioned only in passing is that the Discussion Game is being implemented as a part of the regular curriculum of the classroom. Put simply, this is done to ensure that the students take the activity seriously. One of the potential benefits of the handheld technology relative to standard desktop or laptop computers is that it encourages students to engage physically in the activity, the hope being that physical engagement encourages mental engagement (Klemmer, Hartmann, & Takayama, 2006; Rambusch & Ziemke, 2005). One worry with any novel technology-based curriculum is that students will view it as a mere novelty and as a result will not take it seriously or learn as effectively. On the basis of a number of pilot implementations in various classrooms and school districts, we have found that the students were more focused on the activity when it was integrated with the classroom's current topic of study. Classroom teachers involved in these pilot studies have concurred with this observation. The concern about implementing the Discussion Game in classes was that students would view the activity only as a chance to move around the room, talk with their friends, and play with a new piece of technology. To combat this possibility, we have worked with teachers to use the Discussion Game in the context of a current unit of study as a way to debrief topics of study periodically throughout the school year. Students then become familiar with the activity and view it as a substantive part of the learning process.

Each of the classes involved in the Discussion Game study regularly participates in handheld-based simulations other than the Discussion Game. These other participatory simulations, also developed by the MIT Teacher Education Program, focus on science topics and encourage the students to interact with each other as they use the technology.[6] Because the students are fairly familiar with the handheld computers, we believe they perceive them as less of a novelty.

The regularity of the activity also helps to fight against a common problem associated with experiments. Although controlled environments allow the researchers to isolate particular potential causal mechanisms and to manipulate specific variables while holding others constant, the drawback of the laboratory setting is that subjects may behave differently there than they would in a more natural setting. Experimenter demand effects (Levitt & List, 2007; Orne, 1962) may lead subjects to act as they perceive the experimenter hopes given the design of the experiment. Hawthorne effects (Adair, 1984; Mayo, 1945; Roethlisberger & Dickson, 1939) may show evidence of changes in behavior that are due to generic changes in environments rather than to the specific ones under study. And, more generally, the laboratory setting may be different enough from the real world in important ways that the results of the experiment are not externally valid.

To the extent that we are interested in the actions, behaviors, and learning of children in the classroom, it is important to study children *in the classroom*. Students are perceptive of adults' motives, however. It is therefore important that students do not perceive themselves to be subjects in an experiment and that they do not perceive the activity to be contrived. By making the activity a regular part of the curriculum that recurs throughout the year, and by using it as a pedagogical tool rather than as a break from normal learning activities, we hope to ensure that students behave in this activity as they do in other classroom activities. We view it as a strength that we are able to document patterns of behavior and communication between students in what they view to be the normal course of schooling.

The handheld computers collect data recording the time and names for each meeting between students and of each student's opinion or estimate along with a time stamp. Presently, the only student-entered data we collect at the time of each meeting are the student's answers. Each student is asked to list his or her five closest friends in class and the five students he or she considers the most academically able. From these responses, we are able to construct measures of social networks. We are also able to construct measures of how each student is perceived by her or his peers, both generally and by specific students. Teachers are also given a survey in which they are separately asked to rank or rate each student along two dimensions: academic ability (demonstrated by student performance in the classroom) and social popularity. These two sources of qualitative data—from the students themselves and from the classroom teacher—can be used to create descriptors of each student that add texture to the analysis of student interactions. This rich set of variables might also be supplemented with students' grades and standardized test scores, as well as with demographic (e.g., gender, race, family income) information about students gathered via surveys.

What Questions Can Data From the Discussion Game Answer?

Having described the basic setup of the Discussion Game, we now turn to the set of questions that one might use this activity to answer. These questions fall into two broad categories, and within each there is a large number of interesting variants. Because the focus of this article is to describe the potential of this new methodology, we do not attempt to provide an exhaustive list. Rather, we describe a number of specific questions with the understanding that there are many other possibilities.

Describing Social Networks

First, the Discussion Game is well suited to address questions involving social networks within classrooms. Before studying the effects of social networks and of peer interactions, it seems natural to study the networks themselves. One point we made earlier is that a lot of research on peer effects in social science treats fellow neighborhood residents or fellow students in the classroom as peers. In addition to masking all the interactions between students that are likely to be important determinants of peer effects, this crude description of relevant peers ignores that even within a group as small as a classroom, each student interacts more intensively with a different set of peers.

There is a large literature devoted to describing and analyzing the characteristics of these social networks. Historically, much of

[6] The Virus Game described above is one example. Descriptions of other participatory simulations can be found at the MIT Teacher Education Program website (http://education.mit.edu/pda/games.htm).

this research has been based on self-reports of relationships and interactions (Kossinets & Watts, 2006; Newman, 2003; Wasserman & Faust, 1994). Of particular relevance here is the early work of Rapoport and Horvath (1961), who study friendship networks of students. Survey-based network data have been criticized for their accuracy and subjectivity (Marsden, 1990) and for their static nature (Kossinets & Watts, 2006). Advances in technology have allowed researchers to collect more detailed and dynamic data on social interactions. Examples include the use of cellular telephones (Eagle, 2005; Raento, Oulasvirta, & Eagle, 2007), a network of long-distance telephone calls (Aiello, Chung, & Lu, 2002), and patterns of e-mail communication between students at a university (Kossinets & Watts, 2006).

In the spirit of this recent use of technology, we use the handheld computers in the Discussion Game to collect data on which students choose to pair with whom, and when. As mentioned previously, sometimes students are told by the teacher who to meet with and when on the basis of a randomly generated order. However, at other times, students are simply told to pair with someone and, when the first round is finished, to find someone else. There is interesting information in who students choose to pair with and when. One might also ask how students' pairing choices vary as the domain, goals, and rules of the activity change. Do students choose different partners in the opinion and fact versions of the activity? Do the characteristics and timing of pairings change if students are given explicit incentives on the basis of how close their estimate is to "the truth" or how close their opinion is to the majority opinion? Do students choose to meet with their friends earlier in the opinion domain but with those they perceive as smart in the estimation domain? Do these patterns differ by gender? Do the patterns of pairing change over the course of the school year, and if so, how? How often do girls pair with boys, and which types of girls and boys are likely to pair? How common are interracial pairings? How often do students from different economic strata pair?

One nice feature of using these data to describe social networks is that they are based on actual choices made by students in the classroom environment. The caveat about the external validity of experiments conducted in laboratory settings applies also to observation of nonexperimental behavior in laboratory settings. If we want to learn about how students interact, then it is important to observe these interactions in a setting about which we care.

A second appealing feature of these data is that they are dynamic. As mentioned above, many descriptive measures of social networks are based on point-in-time data or on questionnaires that refer to long-term relationships. Although these are important descriptions of stable social networks, they are not designed to study how these networks evolve. Because we observe student-driven interactions under different circumstances, and at different times, we can describe how the social interactions respond to these changes and how they evolve. In this case, the time frame might be within a single class period or over an entire academic year.

The activity we have described also has important weaknesses for studying social networks. One weakness is that by design only one student can meet with any given student at any time. Many in the class may prefer to meet with a particular student in the first round of pairings. We only observe one of these pairs. Because we do not observe the details of the allocation mechanism—how it is determined that one student gets his or her first choice while others are forced to look elsewhere for a partner—we have no data on the preferences of students toward those with whom they were not able to pair. One might imagine instituting a bidding system, in which students "pay" for the right to meet with a particular student. Subject to caveats about strategic bidding behavior, data on students' bids could be used to infer their preferences over peers. In implementing such a system, we would have to be very mindful of the concerns raised above about measuring behavior in artificial settings.

A second obvious weakness is that the activity is done in a small selected set of classrooms. Ideally, one would like to implement the Discussion Game in a large nationally representative sample of classrooms. In practice, it is necessary to recruit supportive school administrators and teachers who are willing and able to alter their lesson plans throughout the year to integrate this activity. The set of teachers that fits this description and has this freedom is surely not representative. It is important to keep this fact in mind when interpreting results.

Testing Possible Mechanisms Underlying Peer Effects in Classrooms

The second set of questions the Discussion Game is designed to answer relates to the mechanisms underlying peer effects. As we considered earlier, sociologists and economists have produced far more evidence on whether peers affect outcomes than on how this happens. This lack of empirical evidence is not for a lack of theories though. We hope the data we collect from variants of the Discussion Game can shed light on some small subset of these theories.

Suppose a student forms an answer to a Fermi question according to the equation

$$y_{i,t+1} = \alpha y_{j,t} + \gamma y_{i,t} + \boldsymbol{X_i}\beta + \varepsilon_{i,t+1},$$

where $y_{i,t+1}$ denotes student i's estimate after the current round, $y_{j,t}$ denotes the estimate submitted by student i's current-round partner after the previous round (i.e., at the beginning of the current round), $\boldsymbol{X_i}$ is a vector of characteristics of student i, $\varepsilon_{i,t+1}$ is a random error term, and α, β, and γ are parameters to be estimated. The parameter of interest, α, is the effect on student i's estimate of the estimate his or her current partner relays at the beginning of the current round.

If this simplified model properly characterizes each student's thought process, and students are randomly assigned to partners, then we can estimate the effect of a peer's estimate by ordinary least squares regression.[7] This analysis addresses the question, do

[7] Other assumptions are necessary for interpretation of the resulting estimate. The common shock problem described above might reasonably be ignored because every student, and therefore every pairing, is participating in the same classroom at the same time. The reflection problem remains, however. One should recognize that the estimate of α is a reduced-form parameter that includes both the effect of the peer's lagged estimate and the effect of the peers current estimate, to the extent that it is correlated with her or his own lagged estimate (see, e.g., Manski, 1993; Sacerdote, 2001, for discussion of endogenous and exogenous peer effects). One must also be cognizant of the fact that each student enters the sample on both the left- and right-hand sides of the above equation when estimating standard errors.

students move their estimates closer to their peers' after discussing the problem?

Following from this simple relationship, one can ask a number of interesting questions of two related forms. First, how does this relationship depend on the characteristics of the students in the pair? Second, how does it depend on the broadly defined setting within which students operate?

With respect to the first type of question, one might ask whether girls are more likely to move their estimate closer to boys or to girls? Are minority students more strongly influenced by students within the same minority group or by others? Do the answers to these questions depend on whether the class is estimating the answer to a Fermi question or offering opinions on a controversial topic? Are students more likely to be influenced by the academically talented students in the classroom or the popular ones? Are the smart students more influential in the fact-based domain and the popular students more influential in the opinion domain?

The second type of question is closely related to the first. Suppose the teacher was on particular occasions to change the rules of the activity. How do the relationships described above vary with these changes in the rules? The change from the Fermi question to the opinion statement is one example. Another might be the offer of explicit incentives based on how close students' estimates are to the truth. Are students more likely to seek out the help of well-informed students when something tangible is at stake? Which types of students become more influential in this setting? Another dimension that is potentially interesting is that of time. Do students respond to their peers in the same way at the beginning of the class period as they do toward the end? Does this pattern differ between the fact and opinion domains? Do students wish immediately to have an opinion that is close to others but digest the estimates of many peers before using them to form their own estimate?

Although we think these patterns will be enormously interesting no matter what they show, they raise further questions about how peers influence each other. To begin to address some of these questions, we are working to incorporate collections of data on students' rationales. Using a variant of the Discussion Game, Yoon (2008) has studied how *memes*—a term for ideas or thoughts that can replicate like genes—flow through classrooms.[8,9] We hope to build on this work by tracking how specific rationales flow from different types of students to others. Once we can track individual memes, or rationales, we might give hints to randomly selected students and then track how these memes flow throughout the class. For example, suppose the Fermi question is "What is the distance in inches from Minsk, Belarus to Sacramento, California?" The teacher could secretly tell one student that the distance from New York, New York to Chicago, Illinois is 714 miles and tell another student that there are 63,360 in. in a mile. Each of these hints might help the student to narrow down his or her estimate, and the student may in turn tell the hint to peers. Whether the peers use this information depends on how well each successive student explains its value. Do these hints flow throughout the classroom, and if so, how quickly? Which types of students are more likely to pass these memes on to others?

Discussion of the Discussion Game

Although the data collection techniques described are innovative in many ways, they still fall short of work already being conducted by developmental psychologists in several ways. Technology is not a perfect substitute for human observation. Researchers in the classroom can observe characteristics of interactions they did not expect to be of interest. Videotaping allows researchers to examine interactions in detail after the fact to test secondary hypotheses. In comparison, the technology we describe is limited by the preintervention decisions and imagination of the researcher.

There are countervailing benefits to the use of technology to collect data on peer interactions, however. First, technology like the handheld computers reduces the cost of data collection. As a result, the researcher can compile a much larger sample of observations on peer interactions than he or she would be able to if direct observation were necessary. Second, the automation of data collection ensures that subconscious observer biases do not taint the information collected. Researchers who do direct observation work hard to avoid these biases, and though they may be successful, it is reasonable to believe that automation is a preferable method on this count. The presence of the technology itself may directly affect student behavior, of course. Motivated by this concern, we have argued that it is crucial for the technology to be integrated into the classroom setting in such a way that students view its presence as a part of normal classroom activities. That brings us to the third consideration: Handheld technology can be integrated into the classroom more easily than direct observers or videotaping. To the extent that researchers are concerned that children act differently when they know they are being watched, future technology may be even more useful.

Last, we began this article discussing how difficult it is to identify the effect of one child's behaviors or characteristics on another child's behaviors or outcomes. Among the descriptions of detailed data collection and underlying mechanisms, it is important not to lose sight of this core problem. The collection of more detailed data does not obviate the need for a valid research design to identify peer effects. It only makes such a research design more valuable. We hope that the use of technology to automate data collection will make it easier for researchers to implement research designs with random-assignment of peers. Once a teacher has integrated the simulations that use the handheld technology into the classroom, it is a minor step to introduce teacher-dictated pairings or groupings, which could be randomly generated. If our hope is correct, this could be the most valuable attribute of the technology for peer effect research.

We conclude with a brief note about the interdisciplinary nature of this collaboration. Attempts at cross-disciplinary collaborations can often be frustrating and unproductive. Collaborators must find a common language to discuss ideas and avoid the costs associated with teaching others commonly accepted concepts. In this case, however, the differences in our skills proved complementary. As a result, we believe together we produced a research design—the combination of the technology, the curriculum, the activity, and the set of questions—that is far superior to what we could have produced individually. We encourage researchers interested in social interactions among children to attempt similar collabora-

[8] See also Yoon and Woodruff (2003).

[9] The term *meme* was coined by Richard Dawkins (1976) in *The Selfish Gene*.

tions. The promise of new technologies to collect better and better data should be a great incentive for social scientists to do so.

References

Adair, J. G. (1984). The Hawthorne effect: A reconsideration of the methodological artifact. *Journal of Applied Psychology, 78,* 413–432.

Aiello, W., Chung, F., & Lu, L. (2002). Random evolution of massive graphs. In J. Abello, P. Pardalos, & M. Resende (Eds.), *Handbook of massive data sets* (pp. 97–122). Dordrecht, the Netherlands: Kluwer Academic.

Berndt, T. J., Perry, T. B., & Miller, K. E. (1988). Friends' and classmates' interactions on academic tasks. *Journal of Educational Psychology, 80*(4), 505–513.

Boulton, M. J., & Underwood, K. (1992). Bully/victim problems among middle school children. *British Journal of Educational Psychology, 62,* 73–87.

Coie, J. D., Lochman, J. E., Terry, R., & Hyman, C. (1992). Predicting early adolescent disorder from childhood aggression and peer rejection. *Journal of Consulting and Clinical Psychology, 60*(5), 783–792.

Cook, P., & Ludwig, J. (1997). Weighing the burden of "acting White": Are there race differences in attitudes toward education? *Journal of Policy Analysis and Management, 16*(2), 256–278.

Dawkins, R. (1976). *The selfish gene.* New York: Oxford University Press.

DeRosier, M. E., Kupersmidt, J. B., & Patterson, C. J. (1994). Children's academic and behavioral adjustment as a function of the chronicity and proximity of peer rejection. *Child Development, 65*(6), 1799–1813.

Dodge, K. A. (1986). A social information processing model of social competence in children. In M. Perlmutter (Ed.), *The Minnesota Symposium on Child Psychology* (Vol. 18, pp. 77–125). Hillsdale, NJ: Erlbaum.

Duncan, G. J., Boisjoly, J., Kremer, M., Levy, D. M., & Eccles, J. (2005). Peer effects in drug use and sex among college students.*Journal of Abnormal Child Psychology, 33*(3), 375–385.

Eagle, N. (2005). *Machine perception and learning of complex social systems.* Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Fordham, S., & Ogbu, J. (1986). Black students' school success: Coping with the burden of "acting White." *The Urban Review, 18*(3), 176–206.

Fryer, R., & Torelli, P. (2005). *An empirical analysis of "acting White"* (NBER Working Paper No. 11334). Retrieved January 4, 2008 from http://www.nber.org/papers/w11334

Gillies, R. M. (2003). The behaviors, interactions, and perceptions of junior high school students during small-group learning. *Journal of Educational Psychology, 95*(1), 137–147.

Hanish, L. D., Martin, C. L., Fabes, R. A., Leonard, S., & Herzog, M. (2005). Exposure to externalizing peers in early childhood: Homophily and peer contagion processes. *Journal of Abnormal Child Psychology, 33*(3), 267–281.

Hartup, W. (1996). The company they keep: Friendships and their developmental significance. *Child Development, 67*(1), 1–13.

Hoxby, C., & Weingarth, G. (2006). *Taking race out of the equation: School reassignment and the structure of peer effects* [mimeograph]. (Available from Harvard University)..

Hymel, S., Rubin, K. H., Rowden, L., & LeMare, L. (1990). Children's peer relationships: Longitudinal prediction of internalizing and externalizing problems from middle to late childhood. *Child Development, 61*(6), 2004–2021.

Johnson, D., & Johnson, R. (1985). The internal dynamics of cooperative learning groups. In R. Slavin, S. Sharan, S. Kagan, R. Hertz-Lazarowitz, C. Webb, & R. Schmuck (Eds.), *Learning to cooperate, cooperating to learn* (pp. 97–102). New York: Plenum Press.

Kandel, D. (1978). Homophily, selection, and socialization in adolescent friendships. *The American Journal of Sociology, 84*(2), 427–436.

Klemmer, S., Hartmann, B., & Takayama, L. (2006). *How bodies matter: Five themes for interaction design.* In *Proceedings of the Sixth ACM Conference on Designing Interactive Systems* (pp. 140–149). University Park, PA: ACM Press.

Kling, J., Liebman, J., & Katz, L. (2007). Experimental analysis of neighborhood effects. *Econometrica, 75*(1), 83–119.

Klopfer, E., & Yoon, S. (2005). Using palm technology in participatory simulations of complex systems: A new take on ubiquitous and accessible mobile computing. *Journal of Science Education and Technology, 14*(3), 287–295.

Klopfer, E., Yoon, S., & Rivas, L. (2004). Comparative analysis of Palm and wearable computers for participatory simulations. *Journal of Computer Assisted Learning, 20,* 347–359.

Knight, G., & Bohlmeyer, E. (1990). Cooperative learning and achievement: Methods for assessing causal mechanisms. In S. Shlomo (Ed.), *Cooperative learning: Theory and research* (pp. 1–22). Westport, CT: Praeger.

Kochenderfer, B., & Ladd, G. W. (1996). Peer victimization: Cause or consequence of school maladjustment? *Child Development, 67*(4), 1305–1317.

Kossinets, G., & Watts, D. (2006, January 6). Empirical analysis of an evolving network. *Science, 311,* 88–90.

Ladd, G. (1999). Peer relationships and social competence during early and middle childhood. *Annual Review of Psychology, 50,* 333–359.

Ladd, G. W., Kochenderfer, B. J., & Coleman, C. C. (1996). Friendship quality as a predictor of young children's early school adjustment. *Child Development, 67*(3), 1103–1118.

Ladd, G., & Mize, J. (1983). A cognitive-social learning model of social skill training. *Psychology Review, 90*(2), 127–157.

Lazear, E. (2001). Educational production. *Quarterly Journal of Economics, 116*(3), 777–803.

Levitt, S., & List, J. (2007). What do laboratory experiments tell us about the real world? *Journal of Economic Perspectives, 21* (2), 153–174.

Manski, C. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies, 60,* 531–542.

Marsden, P. (1990). Network data and measurement. *Annual Review of Sociology, 16,* 453–463.

Mayo, E. (1945). *The social problems of industrial civilization.* Boston: Harvard University, Graduate School of Business Administration.

McFarland, D. (2001). Student resistance: How the formal and informal organization of classrooms facilitate everyday forms of student defiance. *American Journal of Sociology, 107*(3), 612–678.

Moffitt, R. (2001). Policy interventions, low-level equilibria, and social interactions. In S. Durlauf & P. Young (Eds.), *Social dynamics* (pp. 45–82). Cambridge, MA: MIT Press.

Newman, M. (2003). The structure and function of complex networks. *SIAM Review, 45*(2), 167–256.

Oakes, J. (2005). *Keeping track: How schools structure inequality.* New Haven, CT: Yale University Press.

Orne, M. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17*(11), 776–783.

Parker, J. G., & Asher, S. R. (1993). Friendship and friendship quality in middle childhood: Links with peer group acceptance and feelings of loneliness and social dissatisfaction. *Developmental Psychology, 29*(4), 611–21.

Raento, M., Oulasvirta, A., & Eagle, N. (2007). *Smartphones: An emerging tool for social scientists* [mimeograph]. (Available from Massachusetts Institute of Technology)

Rambusch, J., & Ziemke, T. (2005). The role of embodiment in situated learning. In B. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (1803–1808). Mahwah, NJ: Erlbaum.

Rapoport, A., & Horvath, W. (1961). A study of a large sociogram. *Behavioral Science, 6,* 279–291.

Roethlisberger, F. J., & Dickson, W. J. (1939). *Management and the worker: An account of a research program conducted by the Western Electric Company, Hawthorne Works, Chicago.* Cambridge, MA: Harvard University Press.

Sacerdote, B. (2001). Peer effects with random assignment: Results for Dartmouth roommates. *Quarterly Journal of Economics, 116*(2), 681–704.

Slavin, R. (1983). When does cooperative learning increase student achievement? *Psychological Bulletin, 94*(3), 429–445.

Vandell, D. L., & Hembree, S. E. (1994). Peer social status and friendship: Independent contributors to children's social and academic adjustment. *Merrill-Palmer Quarterly, 40*(4), 461–477.

Vaughn, B., Vollenweider, M., Bost, K., Azria-Evans, M., & Snider, J. (2003). Negative interactions and social competence for preschool children in two samples: Reconsidering the interpretation of aggressive behavior for young children. *Merrill-Palmer Quarterly, 49*(3), 245–278.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications.* New York: Cambridge University Press.

Wilson, W. (1987). *The truly disadvantaged.* Chicago: University of Chicago Press.

Yoon, S. (2008). An evolutionary approach to harnessing complex systems thinking in the science and technology classroom. *International Journal of Science Education, 30*(1), 1–32.

Yoon, S., & Woodruff, E. (2003). *Rethinking teaching and learning about scientific and technological issues: Exploring the development of memes in a complex system.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.

Zimmerman, D. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and Statistics, 85*(1), 9–23.

Received February 19, 2007
Revision received September 8, 2007
Accepted December 20, 2007 ■