

# Toulouse Network for Information Technology

Annual Meeting, October 18-19, 2013  
Microsoft Campus, Redmond, WA

## List of Abstracts

### Jonathan Levin (Stanford)

*The Data Revolution and Economic Analysis* (joint with L. Einav)

Abstract: Many believe that “big data” will transform business, government and other aspects of the economy. In this article we discuss how new data may impact economic policy and economic research. Large-scale administrative datasets and proprietary private sector data can greatly improve the way we measure, track and describe economic activity. They also can enable novel research designs that allow researchers to trace the consequences of different events or policies. We discuss some of the challenges in accessing and making use of these data. We also consider whether the big data predictive modeling tools that have emerged in statistics and computer science may prove useful in economics.

### Daron Acemoglu (MIT)

*Managing Innovation in a Crowd* (joint with M. Mostagir and A. Ozdaglar)

Abstract: Crowdsourcing is an emerging technology where innovation and production are sourced out to the public through an open call. In this paper, we develop a crowdsourcing model that considers two kinds of innovation tasks, structured and unstructured. Structured tasks have unknown difficulties, but these difficulties come from a known distribution. They can be performed by any worker who possesses the necessary skills to complete them. On the other hand, unstructured tasks have no precise measure of difficulty, and the skills required to complete them are not well understood. They are usually completed by a worker through chance, human intuition, and guess work. Whether a task is structured or unstructured is an unobservable attribute in our model. At the center of crowdsourcing is a resource allocation problem: there is an abundance of workers but a scarcity of high skills. This problem is exacerbated by the fact that the type and exact difficulties of innovation tasks may not be known in advance, so tasks that require skilled labor cannot be identified and allocated ahead of time. When the pool of tasks consists of only structured tasks, we show that the solution to the problem takes the form of a skill hierarchy, where tasks are first attempted by low-skilled labor and high- skilled workers only engage with a task if

workers with lesser skills are unable to finish it. If the pool of tasks contains both structured and unstructured tasks, then the optimal solution is generally not a hierarchy.

We next focus on the optimal hierarchies for structured tasks. Organizing these hierarchies in crowdsourcing is difficult because firms have little information about the skills of the workers, and workers may have an incentive to misrepresent their skills. This complicates the firm's problem, since it now wants to find an optimal assignment of workers to tasks even though it knows neither the difficulties of the tasks nor the skills of the workers. We give a dynamic pricing mechanism for tasks that utilizes the concept of self-selection. Each time a task is attempted and not finished, its price (reward upon completion) goes up. By correctly setting the prices, the mechanism provides an incentive for workers to sort themselves into an optimal hierarchy, i.e. workers participate in the same level of the hierarchy that would be produced if the firm had knowledge of the workers' skills, ultimately leading to the desired optimal matching between workers and tasks.

### **Nicholas Bloom (Stanford)**

*IT and Management in America* (joint with E. Brynjolfsson, L. Foster, R. Jarmin, I. Saporta-Eksten and J. Van Reenen)

Abstract: One of the great benefits of the increased use of IT in firms is that it supports the introduction of more structured, data focused, management practices. However, it is not clear whether more structured practices are actually linked to higher productivity – they could for instance cause more bureaucratization. This paper has two goals a) to study the effect of structured practices and b) to study how they are encouraged by IT. The first part has been completed and is presented in a working paper “Management in America” by the same co-authors. The second part is being developed at the present time and the first results will be presented at the TNIT meeting in Seattle. To conduct this study we use a recent survey that the Census Bureau recently conducted a survey of management practices in over 30,000 plants across the US, the first large-scale survey of management in America. Analyzing these data reveals several striking results. First, more structured management practices are tightly linked to better performance: establishments adopting more structured practices for performance monitoring, target setting and incentives enjoy greater productivity and profitability, higher rates of innovation and faster employment growth. Second, there is a substantial dispersion of management practices across the establishments. We find that 18% of establishments have adopted at least 75% of these more structured management practices, while 27% of establishments adopted less than 50% of these. Third, more structured management practices are more likely to be found in establishments that export, who are larger (or are part of bigger firms), and have more educated employees. Establishments in the South and Midwest have more structured practices on average than those in the Northeast and West. Finally, we find adoption of structured management practices has increased between 2005 and 2010 for surviving establishments, particularly for those practices involving data collection and analysis. This seems to corroborate the fact that IT, which should favor structured practices, could have an important role in increasing productivity and profitability.

### **Michael Whinston (MIT)**

*Channel 5 or 500? Vertical Integration, Favoritism, and Discrimination in Multichannel Television*

*Gregory Crawford (University of Zurich), Robin S. Lee (NYU Stern), Breno Vieira (Stanford GSB), Michael Whinston (MIT), Ali Yurukoglu (Stanford GSB)*

Abstract : We analyze the impact of vertical integration between upstream content and downstream distribution firms on carriage, positioning, and viewership of cable television channels. Using fourteen years of comprehensive channel lineup data for the population of U.S. cable systems and individual level viewership data, we examine how vertically integrated cable system operators differentially treat their own versus rival channels. We find that: (i) integrated operators carry their own channels more often than other non-integrated operators, (ii) integrated operators place content that is the same genre as their own on less widely available tiers, (iii) integrated operators place their own channels in lower positions than other operators place the same channels, and (iv) lower channel positions are favorable because they lead to more viewership. We also find weaker evidence that integrated operators place content that is rival to their own content in worse channel positions. However, in other cases, we find that vertically integrated firms treat content that is rival to their own just as other unaffiliated distributors treat that content.

### **Susan Athey (Stanford)**

*The Nature and Incidence of Software Piracy: Evidence from Windows* (joint with S. Stern)

Abstract: This paper uses data from Microsoft Windows 7 to analyze software piracy by individual consumers, which we call "retail" piracy. Our data is available at a fine level of granularity, which allows us to uncover new insights about software piracy. We begin by studying the ways in which software is pirated. We find that a large share of retail piracy occurs by using keys that are posted on internet sites such as PirateBay, and that a small number of keys accounts for a large share of piracy. We then examine how piracy varies with the economic and institutional environment, showing that it responds to GDP and intellectual property protection. We show that piracy responds to economic forces such as price and the time/bandwidth cost of downloading pirated versions. We also show that frictions in the ability to pirate has a non-trivial impact on piracy. For example, we show that when countries disallow connection to PirateBay, piracy declines.

### **Nicolas Lambert (Stanford)**

*Dynamic Belief Elicitation* (joint with C. Chambers)

Abstract: A risk neutral agent has privately held beliefs about the outcome of a random variable, as well as private information as to a dynamic signal structure. Signals are informative about the outcome of the random variable, but are unobservable to the principal. A principal wants to elicit (i) the agent's joint belief

about the random variable and signal structure, and (ii) after each signal realization, the signal which was observed. We show that the principal can elicit these two objects in a strictly incentive compatible way by generalizing the classical notion of a scoring rule: when no signals are observed, our device collapses to a classical scoring rule. If only one signal is observed, the principal first asks the agent to select from a menu of classical scoring rules; after the signal realization, the principal asks the agent her probability according to the scoring rule she selected. The construction is extended recursively when there are more signals; i.e. if two sequential signals are observed, the principal asks the agent to select from a menu of menu of scoring rules. All payoffs to the agent occur after the publicly observed realization of the random variable.

### **Alessandro Bonatti (MIT)**

*Selling Cookies* (joint with D. Bergemann)

Abstract: We develop a model of data pricing and targeted advertising. A monopolistic data provider determines the price to access “cookies,” i.e. informative signals about individual consumers’ preferences. The demand for information is generated by advertisers who seek to tailor their spending to the value of each consumer. We characterize the set of consumers targeted by the advertisers, and the optimal monopoly price of cookies. The ability to influence the composition of the advertisers’ targeted set provides incentives to lower prices. Thus, the monopoly price of data is decreasing in the reach of the database and increasing in the number of competing sellers of exclusive data. Finally, we explore the implications of nonlinear pricing of information, and characterize the exclusive data sales that emerge as part of the optimal mechanism.

### **David Rothschild (Microsoft Research)**

*The Mythical Swing Voter and Non-Representative Polling*

Abstract: The only acceptable form of polling in the multi-billion dollar survey research field utilizes representative samples. We argue that with proper statistical adjustment, non-representative polling can provide accurate predictions, and often in a much more timely and cost-effective fashion. We demonstrate this by applying multilevel regression and post-stratification (MRP) to a 2012 election survey on the Xbox gaming platform. Not only do the transformed top-line projections from this data closely trend standard indicators, but we use the unique nature of the data’s size and panel to answer a meaningful political puzzle. We find that reported swings in public opinion polls are generally not due to actual shifts in vote intention, but rather are the result of temporary periods of relatively low response rates among supporters of the slumping candidate. This work shows great promise for using non-representative polling to measure public opinion and the first product of this new polling technique raises the possibility that decades of large, reported swings in public opinion—including the perennial “convention bounce”— are mere artifacts of sampling bias.

**Matthew Gentzkow (Chicago)**

*Ideology and Online News* (joint with J. Shapiro)

Abstract: News consumption is moving online. If this move fundamentally changes how news is produced and consumed it will have important ramifications for politics. In this paper we formulate a model of the supply and demand of news online that is motivated by descriptive features of online news consumption. We estimate the demand model using a combination of microdata and aggregate moments from a panel of Internet users. We evaluate the fit of the model to key features of the data and use it to compute the predictions of the supply model. We discuss how such a model can inform debates about the effects of the Internet on political polarization and other outcomes of interest.