Innovative Assessment of Collaboration 2014
Panel 5: Statistical Models for Collaboration and Group Dynamics, Closing Remarks and Summary

Alina von Davier

Thank you all for sticking around. This is the last session, but not the least of this conference on analyzing the data and building the research necessary for innovative assessments of collaboration. The perspective we are going to take in this panel is again different, as by now you are or you got used to from panel to panel switching perspectives. This time we are looking at a team performance, but still we are in the realm of statistical models, but this time we focus on group dynamics as opposed to the individuals as you saw in the previous panel. Our panelists this morning are Professor Sy-Miin Chow. She is an associate professor in the Department of Human Development and Family Studies at Penn State University and principal investigator of the Emotions and Dynamic Systems Lab. Our next, the second panelist is Tracy Sweet. She is an assistant professor in the Department of Human Development and Quantitative Methodology at University of Maryland. And our third presenter in principal is Mengxiao Zhu, who is an associate research scientist in the Center for Advanced Psychometrics at Educational Testing Service, a co-organizer of this event. And I mentioned before she unfortunately cannot be here at the meeting and Noshir Contractor, her dissertation supervisor, kindly agreed to give her presentation today. So with this I will ask Sy-Miin Chow to start the panel. Thank you.

Sy-Miin Chow

So I want to start by thanking ETS for inviting me to this conference. So my area of research is in dynamic modeling, including linear and nonlinear dynamical systems models and space-based models. I do have some collaborative work with Jeff Cohn and Daniel Messinger. I'll say more about their work in the context of one study. So first I just want to give you a quick example of the modeling framework that I'm adopting, and this is basically a linear space-based model and what we are trying to do in this context is we have two equations. The second equation basically is the measurement(?) model that tries to relate things that we do see to manifest(?) observe variables to a bunch of latent variables that we don't get to measure directly, and the first equation is basically a dynamic model trying to dictate or show how the latent variables change over time as a function of the latent variables …. T minus 1(?). And we also have covariates X and Z that are related to either the measurement process or the dynamic process. So here I'm showing you a measurement model that gets at continuous data, but it doesn't have to be the case. It can be extended to categorical data, measures that conform to exponential family, binary data and so on and so forth. The dynamic model does not necessarily have to be linear. It can include nonlinear models and so on and so forth.

So this is the first dataset that I'm going to talk about. So this is work in collaboration with Jeff Cohn, a psychologist and computer scientist at University of Pittsburgh and Carnegie Mellon University, as well as Daniel Messinger, psychologist at University of Miami. So here the idea is that they brought in a mom-infant dyad and have them interact with one another, and in the process basically they take automated measurement. So I'll try to steer away from the word non-conscious. So this is automated measurement of information such as lip movement, so curling or curving of the lip in smiling, eye constrictions – when we smile there's constriction in the eyes – mouth opening, and we try to see the evolution of the mom and the infant trajectories over time. So from all these measures we are trying to get a sense of an underlying latent construct representing positivity of the infant and the mom and trying to see the dynamics between the two over time. So there are a few things that we may be able to notice from this kind of data. So for instance we may be able to get a sense of who is the leader of the pack, so who is driving the dynamics. So in some cases the parents may claim that often time it's the baby that is bossing us around, but there may be other times when a parent takes on the role of the leader and these kind of dynamics may not be stationary over time, so this may change over time as well.

So this is to give you an idea of the over-time dynamics of some of the measures that I talk about and there may be covariates that we want to bring in. So one example may be tickling. So in this specific dyad the mom is working hard ticking the baby and in the process getting the baby to smile. This is yet another dyad where the infant is not having such a good time. So the tickling did lead to some surges in positive emotions, but only in the no brow-furrowing phase, so when the baby was doing fine, not complaining, not feeling unhappy.

So we have information that are very time intensive across dyads. So the first step is to think about how we construct dynamic models to represent the interaction and the …. regulation manifested by the dyad, but within the dyad member as well as the connectivity or the coupling between dyad members. And this slide is just to show you briefly what we gain from automated measurements. So to the right we have composite score derived from automated facial analysis and to the left we have human rater expert(?) coding. So you can see that with the automated measurement the rises and the falls tend to be a little bit sharper, so there is a little bit better time position when it comes to capturing the dynamics of the dyads.

So here is one model that we came up with for the first dyad, the one with a lot of tickling going on. So we have two measures for mom, lip upward movements, smiling, smile strength, eye construction. For babies we have three. From there we extracted an underlying construct representing positive asset PA for mom and baby and we found that tickling basically led to elevation in smile, smiling, as well as eye constriction only for the baby, not for the mom. And we do see that by looking at the dyad as a whole that changes the dynamic model that best characterizes each dyad member in turn. So if we were to look at the …. dynamics of each member, then we found that we needed an AR2 model. We need to take into account the life [live?] dynamics at the previous time point as well as the time point before that. But after putting in the coupling of the cross-regression parameter, then we are fine with an AR, a vector(?) AR process.

So here we are seeing that the baby is basically driving the dynamics. There is a cross-regression in fact from baby to mom, but not from mom to baby. But there are also some between-dyad differences. So for this auto dyad for instance there is no significant coupling between the two dyad members and tickling did not lead to significant rises in any of the measures. They only had a marginally-significant effect on increasing smile strength, but only during the phase when the baby is not complaining.

So now I want to switch gears a little bit. The data that I just talked about there, pilot data coming from automated measurements from two dyads only. But oftentimes in collecting this kind of data there are also data coming from human raters who are asked to code very micro-level changes and from there trying to get a set of composite scores that represent the over(?) idea positivity over time. So there are a few things that I want to point out in data like this that we ought to be thinking about in modeling group dynamics. So one is trends, systematic trends over time. In this particular context we have three columns of data. The first one came from a face-to-face experimental setting where the mom and the babies are allowed to interact with one another as they normally would. The middle column is still phase during which the mom was instructed not to respond to the baby at all. And then the third column is the reunion phase during which the mom is then allowed tor reconcile and reconnect with the baby. So with this kind of experimental setup you can image that there are trends going on with babies getting progressively more upset because the mom is not responding to the baby. The trends may not unfold in a very linear or clean-cut way. There may be a downward trend, but it may not necessarily manifest itself in a linear straight way.

On top of the trends though there are also intra-individual variability that we are interested in. So looking at the one-two plot for instance – will I be able to point? Probably not. So in addition to this downward trend there may also be a sense in which how we can observe the dyad to recover once they have predivations(?) from the usual downward trend. So there is overall trend going downhill, but there may be times when the baby is especially upset or the baby is especially positive, and that may be seen as a predivation from the usual trajectory. And once we have that kind of predivation it would be interesting to see how the dyad members recover from the predivation and how they are coupled to one another. So in time series analysis oftentimes we want to be able to detransfer(?) before studying these kind of patterns of intra-individual variability.

But of course in time series now there is a lot of times we are dealing with data where you want to be able to tease apart and remove trends due to global warming and inflation or other things before you look at intra-individual dynamics. But in dyadic interaction though we want to keep the trend in because we know that a lot of time the trends basically reflect co-regulation of the dyad, so we want to be able to get at the trend and we want to be able to get at intra-individual variability.

So this is one example of a possible model to get at that. Notice that in this plot the circles and squares all represent very different things, so I just want to try to highlight what they mean conceptually. So the idea here is that we have some measurements of mom and some measurements of infant and we are trying to extract two components, one representing a trend component, the ….. The …. basically we are trying to get at a trend using a random watt(?) model. One of the ways in which people have described this kind of model would be imagining a drunk, someone who's drunk trying to take a step randomly, either to the left or to the right, and then trying to see the position of that person over time. I can reassure you that that did not happen to any of the speakers or ETS employees last night. So that is how we get at a trend. So this is how we get at the plot at the top where there is this kind of downward trend. That is not necessarily unfolding(?) in a linear way. And we have the alpha there trying to get at intra-individual variability as well as inter-dyad connection using a …. AR model.

So and there were some discussions about how we may incorporate team-level outcome measures in this type of a model. So in the context of this model one possibility may be to have an outcome measure derived from a slower timescale. So for instance having expert coding or other measure reflecting the quality of the interaction as a separate outcome measure and trying to ask ourselves what are the components that would predict outcome? Is it the trend or is it the intra-individual variability? Is it the alpha or is it the …. But this is just one example of some of the ways in which we can go about integrating outcome measures and response or dependent variables reflecting the performance of the team or the dyad.

And there are a lot of other issues too that may be at work when we study dynamics over time. The dynamics may not be constant over time. So in this specific paper we look at the association between mom and baby, not necessarily imposing directionality, and seeing how that regression causation changes over time in the context of each of the experimental figures. So we do see for instance that during the still phase we don't see time-varying changes in that regression, regression association over time. But in some sections such as during a reunion section there seems to be more volatility in how the association between mom and baby unfolds over time.

To pave the stage for the next speaker, there are also other possibilities. So if we want to move beyond dyad to impose or to get to larger framework where we have multiple group members, there are also methods that allow us to get at that and people have a lot of ideas or concepts about different kinds of network and what may be more efficient in propagating signals or messages over time. And this is collaborative work with Fu(?) Xiang(?) Chow(?) and Amelia Farrow(?) at UC Davis. So in this specific setup data we have couple data over time, and so there are two sets of numbers associated with each node representing the ratings of …. affect of husband and wife.

And the network in this case basically reflects the transitions of the couple as a whole across different nodes …. a context of I think …. …. And you can see that there are different segments or different phases that the couple may be undergoing. So in the lower-left corner for instance there is a phase where both dyad members manifest very high …. affect in synchrony with one another. So one of the possibilities that this kind of dynamic process data open up is asking us or prompting us to think about what are the ways in which the dynamics change over the course of the measurement. So there may be a phase where the couple seems to be doing well. There may be a high-stress phase in which they have high synchrony and …. affect and then there may be a recovery phase. So there are a lot of models that can be used to capture this kind of between-phase differences and the phases may actually unfold within the dyad members.

And finally another idea that I want to flesh out is Losado(?) in the context of mulling(?) team performance used a benchmark …. dynamical(?) systems model called the Lorenz(?) Equation in describing the team performance from different teams. So from the left to the right we have teams that are not flourishing, not doing well, low-performing team to mid-performing teams and high-performing team. So what he is trying to say there is in high-performing team oftentimes they are resilient and you will see this kind of high flexibility in adapting to one another. They came under a lot of attacks because they are basically using a much too-complicated model to describe something that could be captured using a simpler model, but I think the idea there is very interesting. There are ways to fit models like this. I've done work on using …. …. filter or other …. filter to fit this kind of model, but there are a lot of difficulties there when it comes to data like this. There may not be session data. So again if there are easier models or if there are a concept that we can capture using some reframing or reformulation of the model, we may want to go there.

So this is just to summarize some of the points that I mentioned there. So basically my focus here is talking about trends, inter-individual variability, some of the problems and issues that may arise in the context of modeling the dynamics of groups or dyads. I'm just going to stop here.

AD: Our next speaker is Tracy Sweet from the University of Maryland.

Tracy Sweet

Hello. Hopefully everyone can see me over the podium. I can't see my own slides though, but that might be a little bit problematic so I might end up looking to the left and right. But my name is Tracy and I'll be talking about social network models today. So I want to talk a little bit about the motivation for the work that I do and I tend to work on networks of teachers or networks of students. So this gathering we're talking about networks of teams and so the data may not exactly translate, but I hope the models that I present will be useful and interesting and maybe will be a nice foundation for future development in this area. So just these two examples, I want to talk about how we could estimate covaritive(?) effects [facts?] at the node level or at the individual level, and then I want to switch gears and talk about how we might estimate covaritive effects [facts?] at the network level or at the team level. And I'm going to present two different models just to give an idea of kind of how broad these models can be and the types of questions that they can answer.

So the motivation for my work, and my work deals with looking at multilevel models, so looking at modeling networks across, modeling multiple networks. So, for example, if we're thinking about teachers and schools, looking at multiple networks of teachers across various schools, looking at networks of students across various classrooms or across various schools. And so just to give an example, I don't have a slide up here to introduce you to a social network, but suffice it to say that the nodes in this picture are represented by the vertices of the graph and the relationships among them are shown by the edges. I'm going to be talking about binary networks, but you can imagine that all of these models can scale up to either ordinal or continuous tie data.

So the framework that I'm going to introduce is really motivated by this question of modeling these multiple networks. So a lot of the focus in social networks research really is on the single network, looking at very fine-coarse-grain-sized particular aspects of a single network. And so the question is, well, how can we look across multiple networks and answer interesting questions at a coarser grain size? So visually this is the hierarchal network modeling framework and basically we are fitting single network statistical models to each of the networks, and the hierarchy comes in by requiring or assuming that the parameters from all of these single network models come from some common distribution.

So that's the visual for it. The Greek looks something like this. And in the Greek you can see that this class of models is actually pretty flexible. The first thing you'll notice is that we're talking about conditionally-independent models, so oftentimes we might actually model the networks themselves as completely independent. We might model some other dependence assumption. But these models are pretty broad in what types of dependence assumptions we can make across them. We also can incorporate additional levels of hierarchy as needed. So as an example I'm going to talk about a hierarchal latent space model. So the two models that I'm going to talk about today are something called conditionally-independent models, and that's something that's a little bit different from the exponential random graph models that are going to be in Mengxiao's/Noshir's talk.

So the idea behind these models is that we're actually assuming that the ties are independent conditioned on other parameters in the model. So, for example, on the latent space model we make the assumption that all of the individuals in the network have a position in a latent social space. And the idea is that people that are really close together in this social space are much more likely to have a tie than people that are very far apart in this latent social space. And so at the top you can see a kind of a toy example. Imagine you have six people in your network. You can see we have a two-dimensional latent social space, individuals one and two are close, three and four are close and five and six are close. And so then if you look at the network that's generated from these positions we see in fact that there are ties between one and two, three and four and five and six.

In terms of a hierarchal latent space model for binary data, it looks an awful lot like a logistic regression model and that's one of the nice things about these models is they look like your typical linear models. So we see that we can include random intercept to model kind of an overall probability of a tie. We can incorporate covariant and then what I'm doing is I'm tacking on some distance metric. In this example I'm using Euclidean distance, but looking at the distance between those latent space positions. So if the distance is really large I'm subtracting off a lot of that …. probability. If the distances are really small then I'm subtracting off a small amount from the …. probability of a type.

So we can fit this to network data. In this case we have networks of teachers and the relationships among them are advice seeking. So the teachers are asked, "To whom do you go to for advice?" In this particular example I'm just going to talk about a single covariate, and this is where I'm talking about no level as opposed to network level. So this covariance actually is actually at the pair level. It's whether or not two teachers teach the same grade. So if we look at the effects from …. this model, there's a lot of information on this particular slide, and so what you're looking at, these models are all fit using MCMC, so the results that we get are actually …. of a posterior distribution. For those of you that don't like to think in terms of posterior distributions, you could think of them as just colorful 95 percent confidence intervals because they're kind of similar to that. And so what you're seeing in the top graph is what happens when we fit these single network latent space models to the effect of teaching the same grade, and so you see in some instances we have kind of we have kind of egregiously large standard errors. And then the bottom slide, I'm sorry, the bottom plot just shows you the posterior distributions for each of the schools' estimates of teaching the same grade when we set the hierarchal latent space model. And I show this just as an example to see the benefits of fitting a hierarchal model as opposed to fitting separate models for all of the networks, that we're able to actually borrow information across networks. The other nice thing about fitting a hierarchal latent space model is that we are actually able to estimate an overall mean and an overall variance for this particular parameter. So we do see that overall there is a positive effect of teaching the same grade and we would expect that considering these are mostly elementary schools.

So now I'm going to switch gears, so in addition to talking about network level covariance I'm also going to switch gears and talk about a completely different model. And so this model is a type of block model, so again we're still talking about conditionally-independent ties, but now the assumption is that everybody in the network belongs to a group or a block, and people that are in the same group are more likely to have a tie than people that are in different groups. So a typical block model assumes that everybody belongs to a single group, and so we think about when you might see this type of network structure. You could think about departments in a high school, for example, and you have the English teachers, and the science teachers and the math teachers and everybody belongs to their own department. Right? Well, I'm going to talk about mixed memberships to …. block models today. And the difference between a typical block model and mixed-membership block model is that now people are allowed to vary and they're allowed to belong to multiple groups, so each person has their own individual probability vector belonging to any one of those departments.

So I might be a special ed teacher, but work in the Math Department, so I kind of belong to the Special Ed Department as well as the Math Department. If you're thinking about a team, I might be someone that is the team leader, so I might work closely with my own team, but because I'm a leader I might actually then work across teams as well. So just to give an example what do networks look like from block models? Well, if you look at networks we see that we have this subgroup structure. So you have these cliques and you see very few ties between cliques, but very many ties between cliques or blocks rather. When we have a mixed-membership model there's a lot of softening of that. It's a little bit hard to distinguish the actual blocks, and if you can distinguish the actual blocks there's a lot of within block ties.

Here is the model. It's nicely presented as a hierarchal base model. I'm not going to talk about the model, but if anybody has questions we can talk about it after. But I really am going to focus on the theta(?) parameter, and so that's the probability factor that each person has. So each person has their own theta I and that determines the probability of belonging to all of these different blocks. And that's distributed using a D…. distribution with two hyper-parameters, C and gamma. And so I'm going to talk about gamma today. And so when gamma is small we're going to see really extreme membership probability, so you can think about this kind of on a simplex. Right? So we're going to see really extreme probability vectors meaning everybody is going to be really likely to belong to one block and very unlikely to belong to any of the other blocks. When gamma is large what we see is that people are much more likely to belong to multiple blocks. Visually what that looks like is this. So when gamma is small, notice that gamma here is .09, we see that block structure. Right? We see people that have these, you know, we're working with three blocks here. You see that people are much more likely to have a tie within their own block and there are very few across-block ties, whereas when gamma is .6 it's really hard to even tell that there are three blocks. It looks like there's kind of this one big happy family picture.

So then that begs the question, well, networks vary based on how insular or how segregated their subgroups are. There are some networks where the subgroups are quite segregated. There are some networks where the subgroups are so integrated that you can't even tell that there are subgroups anymore. So then the next question was, well, can we relate this to something that's happening at the network level? And so I think this example might go back to the team's work because is there something about the team that's causing these patterns of interactions to appear in this way? So if you look at network one we see really, really segregated blocks, but then when we look at number six again we can't even tell that there are different blocks.

So the model that I introduced has a really, really long name, but basically what we're talking about is a mixed-membership …. block model where instead of having this gamma parameter we're incorporating a linear component for that. So now we can include these network-level covariates that will describe, well, how much segregation is going on between these blocks? And I'm using a dataset, this is fifth-grade network data, friendship network data, excuse me. And so the idea is that if teachers are kind of micro-managey about the ways in which their students interact we might expect to see different types of friendships emerge within these fifth-grade classrooms. And in fact, that's what happens is that teachers that have higher values of what we're calling cognitive facilitation, which means that it's a construct that was observed during a classroom observation, but teachers that have higher levels of this value tend to have students with less segregated classrooms. And just to give an example, you can see when cognitive facilitation scores are small we see these segregated subgroups. When they're larger then we see less segregation, more integrated subgroups. I'm going to stop now because I'm out of time, but I can just put a little bit of information there.

There's a website on this work and talking a little bit about some of the future work. Thank you.

Noshir Contractor

So it's truly a delight for me to channel Mengxiao here today. I'm at that stage now where increasingly I go to places and give talks, and I mention the names of my former graduate students, and people will come up to me later on and say, "You know Mengxiao?" or, "You know Brian?" And it's such a thrill for me to know that my claim to fame now is determined by the illustrious accomplishment of my former graduate advisees. So I'm delighted to be here to channel Mengxiao. This is stuff that she developed for her dissertation, but has been working on it even beyond that, and you'll see towards the end of the slides there is stuff that we'll talk about that we're working on now that is actually a manuscript that is under revision and review at Netbook Science that I'll talk about, at Social Networks, I'm sorry, that I'll talk about.

So the general idea here is motivated by the fact that something that I touched on briefly yesterday in the discussion section, if you were at the opening panel, and that is a lot of the work that is done on collaboration and teams is premised on the fact that the team is already in place and your goal is now to see how to make that team more effective. What Mengxiao did for her dissertation was focus instead on what are the parameters that focus on the assembly of the team in the first place? And a good example of the motivation for this comes from an article that was written by David Ferrucci in the *New York Times* where he; David Ferrucci was the person who was in charge of the Watson Project at IBM, which you may recall was the project that helped to develop this incredible software that could answer any question faster, or ask any question I guess because it was on Jeopardy, but ask any question faster and more accurately than any of the other participants on Jeopardy. And David Ferrucci said that despite the fact that the technology behind was phenomenal, building the team that built Watson was even more challenging than the technology associated with Watson, and that as you can see was the title of this article that he wrote in the *New York Times* at the time.

So the motivation, I had an opportunity actually to talk to David about this, and he gives all kind of interesting stories about how when you come up with a crazy idea, and you want to go form a team and you want to get bright people, then of course all of them are already taken and working on other projects. And then when you get one really good person, really bright smart person to be on the team, and then you go talk to somebody else and say, “Well, I’ve already got; I mean I want you to join this team,” and they go, “Who else do you have?” and I go, “I have person A,” and they go, “Oh, I don’t want to work with person A. Life is too short. He may be smart, but it’s not a person I want to work with.” So there are all kinds of interesting challenges that happen in this area.

So what Mengxiao looked at in particular was focusing on a particular category of teams, and these are project teams which are defined as teams that are time limited, that draw members from different disciplines and functional units and produce one-time output. So in a sense they are one-off teams, if you may, and this was a definition that was based on work by Cohen and Diane Bailey who refer to them as project teams. And so obviously they built on a common set of interests or foresight(?). They often are volunteering, so these self-assembled teams. They often are across geographical boundaries, and cultural boundaries and organizational boundaries. As I said they have a relatively short lifespan and they are often embedded in a social context, so these teams arise from other social ties that might have existed before that.

So what are examples of self-assembled teams? Clearly the kinds of scientific collaborations that we have in academia and in science are examples of self-assembled project teams. Open Source software is another example of these. And the one we're going to talk a little bit more about today and perhaps the less well known are teams that play massively multiplayer online role-playing games. So for those of you not familiar with MMOs or MMORPGs, these are really large games where thousands, millions of people are logged in at any one particular point in time. They take on the role of an avatar and they go kill monsters. That's sort of their main goal. And often they can't do it by themselves, so they join others in teams to do it. Now you might think this is a trivial example, but in fact if you look at the it, the total amount of money involved in these MMOs often goes up to make it the eighth largest country in the world in terms of GDP, so this is not a trivial set of activities that are going on.

So one of the things, the challenges and the opportunities in studying these self-assembled project teams is that we have traditional theory. They formed four buckets – theories, data, methods and computational infrastructure, but if you look at theories of traditional teams with fixed boundaries versus dynamics and self-assembled project teams. So most of our theories have been on fixed teams while these are self-assembled teams that are organizing all the time on their own for a one-off basis. In the past there have been challenges in terms of getting the data to be able to look at this and the fine-grain data, especially data on looking at the team before it was assembled to predict the assembly of the team. Most of what we've been talking about for the last couple of days is collaborations after they were put in place. This is trying to understand if you can get data before the teams are put into place. Of course that's changing now with big data as we'll talk about in a minute.

The third is the method, so we already heard about some brilliant innovative methods from our two previous speakers here, and that's something that is happening very recently, because if you look at network approaches in particular, because of the dependence issues a lot of the traditional approaches cannot be used, and so now we are looking at different ways of being able to bring inferential techniques to go above the descriptive approaches that have traditionally been used in network analysis. And finally, Mengxiao knows this very well, but you can have all the data, you can have all the theories, you can have all the methods, but if you don't have enough computational infrastructure then you're going to spend weeks at a time to run a single model. So she knows this first hand better than almost anyone else in terms of the importance of computational infrastructure to make this happen.

So there were three teams, three approaches that we looked at in an article that we published last year in Social Networks. You could think of a team as a set of individuals. So that's a team as a unit where you look at the individual attributes of the people. You can also look at a team as not just a set of individuals, but a set of individuals and the relationships that they have with each other. What were their priorities? How did that influence it? And then finally you can look at a team, and this is focus of Mengxiao's dissertation, look at a team to see to what extent does the team match with the task. So this is a bipartite way of looking at the network where the red nodes are the individuals and the black square is the task, and the goal here then is to see to what extent is an individual going to be attracted to a particular task and how is that going to be influenced by other individuals who may be associated with that particular task.

So there were several hypotheses in Mengxiao's dissertation. The first was coming from different theoretical backgrounds, the first was self-interest. Low-skilled individuals are more likely to join than high-skilled individuals. The idea is that if you don't have many skills you feel the need to go join a team because that will compensate for your lack of skills. The second is individuals are more likely to join teams on more difficult projects, so that even if you are irrespective of your skill, if the project is more difficult then you're more likely to join it, so you can think of ways of measuring the complexity of the project. The third is that individuals are less likely to join teams that last for a long duration because that's going to require a higher level of coordination cost, and so you're more likely to join a team if you know it's a short-term project rather than a long-term project. You can look at exchange and dependency theory to say individuals are less likely to collaborate with others of the same expertise because you don't need to rely on someone else who has exactly the same expertise as you do, so you're in fact more likely to look at people, to join teams with people who have expertise that complements the expertise you have. On the other hand homophony(?) or birds of a feather does come in place if you're looking to collaborate with others who may be of the same gender, may be of the same age and the same organizational affiliation. And then an individual is also more likely to collaborate with others at the same skill level, so it could be in different areas of expertise but the same skill level means, you know, from an academic point of view you can say, “Are they graduate students? Are they junior faculty or senior faculty?” And then finally the last one is from a co-evolutionary point of view you’re more likely to rejoin a team with somebody who you had worked with previously. So these are different sort of theoretical mechanisms that might explain why you choose to join a team with certain other people.

Now the nice thing about looking at this from a network point of view is that each of these theoretical mechanism that I described previously Mengxiao was able to show that each of them can actually represented as a distinct structural signature. And by structural signature it means that you look at certain kinds of ties that are more likely to occur in this bipartite craft(?) than you would expect by random chance, etcetera. So there are different structural signatures that if these signatures are more likely to be found in the bipartite craft than you would expect by random chance, then you could say that in fact the teaming activity was based on individual skill level or project difficulty or project duration, etcetera.

And so what the dataset that Mengxiao used here was a dataset, as I mentioned, from a massively multiplayer online game. This particular dataset came as a service side(?) log data from Sony Online Entertainment that makes the game called EverQuest 2. You could see that this example on the right-hand side here is a monster and there’s a team of people, characters. These are the people who get on the characters who play the game, and they are teaming together to kill that monster. So that’s typical of what happens within these contexts. Now it’s interesting that this data is actually just a very short slice of data. It was taken from about seven days on one server in North America.

You see the distribution of players on that. What is interesting about the distribution of players, so the black and the red represent males and females. But what is also interesting about this distribution of players is that the centroid of the population of people who play this game is within ten kilometers of the centroid of the population of the United States based on census data. So it's quite representative of the US population in terms of people who are playing this game.

So one of the things that happens is that it turns out that when you work with secondary you can't expect everything. The folks who keep the server data are not keeping it to help researchers. They're keeping it to do logs. It turns out they don't actually identify who are on the teams together. So we had to do sleuthing to see how we would stitch these teams together based upon other log activity, so that if there were a bunch of people who were collocated at a particular place and at a particular point in time, all of them got some incremental points. We used a bunch of these kind of heuristics to say, "Aha, these people must be in a team together." So Mengxiao worked through some really interesting algorithms to identify. So for example from a log like that you could see that player 1 and 2 were in team 1, and players 3, 4, 5, 6 and 7 were in team 2. Also because it was so large we looked at different regions and we did sampling across that, so in some ways that's similar to our first speaker where you look at multiple kinds of data or in multiple networks and then try to aggregate across these networks. So what happens here is you have a very large number of project teams. We have more project teams than players because these players were often involved in multiple teams, etcetera.

So again the analysis method, and Tracy mentioned this briefly, we're talking about exponential random graph models here so this is a class(?) of …. models that essentially says that the probability of one realization of the network on the left-hand side of that equation can be seen as a function of a set of structural signatures or statistic, or what they call sometimes the statistics of patterns that happen within the network so that you can count the number of times a particular signature happens and see to what extent do those statistical signatures inform the probability of one particular realization of the network happening. So again here you are estimating the thetas(?) where each theta is a vector associated with different, the vector of structural signatures which is G in this particular case.

And so what we find here is that it turns out that in terms of the signatures individual skill level is not important. So people who are having a high level, who have low levels of skills are not more likely to join teams, but they are more likely to, people are more likely to join teams when the project is difficult. They are less likely to join teams when the project is of a long duration. They are less likely to join teams with people who have the same expertise as them. They are more likely to join teams with people who have the same age and affiliation, but not gender, and they are more likely to join teams of people who have similar levels of skill similarity, but they're not more likely to join teams with people they have previously been working with.

So again I've summarized these findings here. I am out of time, so I'm going to just touch briefly on some more recent work that Mengxiao has been involved in and this is saying that one of the challenges of exponential random graph models is that there are so many structural signatures that sometimes it becomes difficult to identify which particular signatures are important in this case. So one of the things that Mengxiao along with Stanley Wasserman and a colleague in Russia, Valya(?) Kuskova(?) and I have been working on is to go back to correspondence analysis, which is an older technique, but look at it as an exploratory technique in order to be able to identify what might be possible effect sizes, etcetera, that are important to look at in terms of how teams can be mapped onto, how certain network structures in teams can be mapped onto performance. And so this is some of the more recent work that we've been doing and, for example, we find here that the diversity of a guild in a team, so if teams come from people from different guilds or from different organizations that it seems that that is positively related to short-term performance, so having people from different organizations on your team helps with short-term performance, but hurts long-term performance. So these are still very preliminary findings that we are trying to see how we can use correspondence analysis as an exploratory technique in order to then winnow down the variables that we put into an exponential random graph model. Obviously there are boundary conditions.

Some people are going to say that, "Okay, you have told us something about online games. What does it have to do with the kinds of teams that ETS and others may be more interested in?" I'm going to make the argument, and I think Mengxiao would agree with this, that actually there are quite some analogies between playing online games where you go kill monsters and some of the kinds of other teams that we are more interested in in the offline world, but that remains an open question to empirically verify. And of course now in addition to using bipartite networks, one of the things that we've been working on is something called hypergraphs where we look at hyper-edges where we recognize that people belong to multiple teams, and because we belong to multiple teams and that teams have overlapping membership that the assembly of a team is often influenced by what happens not only within prior relations amongst people in the team, but their ties to other teams and those people's ties to still other teams because information trickles through these and comes to the team. I am way out of time, so I'm going to stop there and thank you again.

AD: = Alina von Davier
SC: = Sy-Miin Chow
TS: = Tracy Sweet
NC: = Noshir Contractor
R: = Other Speakers

AD: So as in the previous panels we'll have now the questions posted for the panelists. So I will read first each of the five questions and then we will see which of the panelists wants to tackle which of the questions. How can we link team dynamic to team or individual outcomes? So we heard this type of question previously in different, from different angles. How can we improve the scalability of the network to large datasets? What are the options for the multilevel analysis to consider both team and individual outcomes? Do we see patterns in the dynamic interactions? Can we improve the model estimation? How can we speed it up? And actually I will probably get, as a moderator I will get back to this question because it's also quite important for the models we discussed previously. The issue of estimating models like this is not, are not trivial, and many of us on the psychometric and statistical side need to put still a lot of work to make things like that actually happen in the real world. So these are the five questions and if any of you, the panelists, wants to start. Perhaps the first, Sy-Minn?

SC: I'll jump in and try to say something about the first one. So it seems that a lot of times it's impossible in terms of, from a construct standpoint to separate the response or dependent variable from the outcome measures, but to me we ought to be thinking about at what timescale the outcomes are unfolding over. So as you mentioned there may be shorter-term outcomes, and so there may be team dynamics or team parameters that may be …. for shorter-term outcomes, but not necessarily in a longer-term outcome way. So even though we may use the same dependent variable as a way to index outcome and we may not have to restrict ourselves to just the same measures, but we can think about how we can accumulate measures over slightly different timescales in thinking about outcomes, and going from there and also drawing(?) measures from other sources based on different variables altogether as an index of outcome.

NC: I'll just add briefly to that. There's a fascinating article that was published in *Science* a couple of years ago by Carter Butts, where he – B-u-t-t-s- where he specifically focuses on recognizing that dynamics need to take into account different timescales. And I think in the previous session we heard some discussion about fractal(?) processes, etcetera, and I think that one of the things that Carter points out is that looking at network dynamics and the impacts on outcomes for example, we need to be able to look at it at multiple timescales that vary from very short-term to very long-term processes, etcetera, and that just looking at it at a single scale, a single level, is often very shortsighted.

AD: If I may I would like to ask a follow-up question. So what do you think that means in terms of assessment? Because as my colleague John Goncala(?) said you cannot bring in people and say, “You are going to be tested for two months in our lab.” So what does it mean from an assessment perspective to think in terms of timescale, short term versus long term? What are your thoughts on that?

SC: So I think even within the context of a single experiment or a single lab section there are oftentimes multiple timescales evolving over time as well. So there are the raw binary data, the zeros and the ones over time, but a lot of times in talking with empirical researchers they may tell you that some of the construct that you’re looking at such as self-regulation or emotion, they don’t change at the second-by-second level. You may be looking at something that’s half a minute to half a minute. So from there, there is already multiple levels or layers of timescales to think about. So if you have the rich time…. to begin with, then you can then start to think about what is level or preferred(?) level of aggregation that would allow you to best predict outcomes. And there are also designs where it allows people to bring in multiple cohorts or repeated measures of multiple lab sessions [sections?] of space, you know, one month apart, two months apart, and so on and so forth, depending on the constructs of interest.

AD: Do you want to pick up on any of this?

TS: No, I think it just also really depends on the goal. Right? So I mean the wonderful thing about having all of this time …. is that we basically could model the process from start to finish. Right? And so the question is, well, how do we aggregate that data? And I think the answer is it depends on what we’re interested in measuring and what our goal is in measuring it. And so I kind of think that addresses a lot of these questions because if we want to look at something that’s happening over the course of ten seconds or thirty seconds, then we need to look at the changes and the interactions that are happening at those times. But if that turns out to be not useful and instead what is useful is looking at the patterns of interactions or the knowledge-sharing that's occurring among the participants at the end or over the course of the interactions, then that's where we want to aggregate the data.

AD: What do you think in terms of the scalability? Actually there was someone in the audience who asked that question for the previous panel and I think the question is valid for the type of methodologies that you presented. Imagine that ETS, for example, administers the TOEFL test every week to about 10,000 people. Now imagine that this 10,000 people from all over the world will be assigned to teams, perhaps even more than one collaborative task, and they will be asked to speak English with each other and we will have this possibility to apply some of these tools. Let's assume that that scenario is possible. How, what do you think that are the challenges that we need to deal with? Can we do that? Besides the technology of course. Let's assume that that's doable. What does it mean to have a sample like that? What does it mean to have an administration like that in terms of the methodology that you present?

NC: I'll start on this one. So that has definitely, has been a huge challenge at least in terms of exponential …. graph models that they don't scale well if you look at the large network. So be careful what you ask for. We get these really large massively multiplayer online games which have networks that are of millions of nodes and we want to look for patterns of structural signatures across them. And the bottom line is that even today if you go to a network that's larger than about seven to 8,000, then it becomes computationally intractable to run. But most of what we're doing out here is the same kind of MCMC estimation that we've talked about earlier and they just are not scalable at that level for the most part. And this is despite the fact that we are using supercomputing infrastructure to be able to run these things. So more recently what a group at the University of Melbourne has been working on is trying to see if we can get robust parameters by sampling these networks. So there's a snowball sampling strategy where you essentially start with random nodes in the network and you go out to say three or four degrees of separation. And you run multiple samples of this, and then you see the extent to which you find some way of either meta-analyzing the parameter estimates that you get from across this to see if in fact there are some robust estimates that you can get across the larger network. We have done this in collaboration with Gary Robbins and one of his postdocs, Alex Duvalla(?). We have run this on social network data that we have from IBM and it was sufficiently, the network was sufficiently large that we could use the snowball sampling, but it was sufficiently small where we could actually run it on the entire thing. And so we find that in fact some of these estimates that we get through using the package that they have developed, it's called SP(?) net for sampling, snowball sampling, P-net(?), that that actually is a promising way of dealing with the scalability issue in that context.

AD: Comments on that Sy-Miin?

SC: I think obviously computation and estimation are always an issue that we have to grapple with and oftentimes we just have to decide on where to put the money. So there are certainly things that you have to give up on, so with a lot of people, then with all these multiple nodes, then at one point it may become computationally not feasible. Same thing though. There you may want to think about perhaps you want to bypass the overtime dynamics and focus on covariance that would predict the network characteristics. And from my standpoint if the interest is in modeling the dynamics, then perhaps there are things that I can simplify or screen out at the item level to reduce the construct there, to try to weed out team members that I think may not be contributing to the overall performance or dynamics of the team. In the studies of human dynamics, for instance, people are often interested in looking at the dynamics among family members, so there are parents, the child or children, relatives. And from there we have a thing about what is the nuclear family that's the unit of interest? What do we want to include? Are there family, extended family members that may be contributing to the family dynamics? So we have to make a decision there and there are certainly ways to simplify things depending on what is of interest and what is important there.

TS: I just want add so we're talking about two different scaling up I guess. So on one hand we're thinking about the enormous amount of process data, the enormous amount of time-stamped data that we have. On the other hand we're talking about these pretty massive networks. And in terms of assessment of teams, the network issue is actually not a big issue because networks, I'm sorry, because the teams that we're thinking about assessing, we're not thinking about having all 10,000 test takers interact in a single network. So that actually makes that problem a little bit less of a problem because we're thinking about dyads, we're thinking about teams of three, four, five, maybe six people. So there's only so many combinations of interactions that you can have given a particular amount of time. So I think the bigger issue for assessment is thinking about the scalability not so much of the networks, but of the time-stamped data. And so then again it goes back to if you have 10,000 students maybe you don't want to be applying really sophisticated models to all 10,000. Maybe the first step is to do kind of a field test to see, well, what things can we explore with this sample of 100 students and then scale up? And so then it becomes just like any other type of assessment problem where you start with a small sample of people and figure out, well, what things are relevant for the models that you want to fit, and then figure out ways to scale up from there.

AD: So if we move on to the next question, so can we perhaps discuss a bit more about multimodal, multilevel analysis at team and individual outcomes?

SC: That's certainly a possibility. So the type of model that I usually work on, we have the dyad members at the, over time at the lowest level, and then there are between-dyad differences at the second level, and then bringing in covariates to try to predict some of those between-dyad differences. So at least in the kind of model that I'm working with that part is more manageable compared to some more complex models involving more members for instance.

AD: Any of you want to?

TS: Yes. I just want to talk about some of the methods that are out there that exist for this already. So thinking in terms of networks, not so much for predicting necessarily team and individual outcomes, but predicting team and individual level covariates. So there are a number of influence models that exist in the networks literature. So thinking about how someone's interactions with someone would actually then influence if we're thinking about in terms of assessment maybe the amount of knowledge that they have or the way that they're approaching a particular problem. And so there are these dynamic network models that actually can do that, so you have this idea that people come together or form these interactions because they have similarities, but because, but then as a result of these interactions people's beliefs, the way that they're thinking, other things start to change and so you kind of have this co-evolving network. And so there are models out there that can estimate these effects both at the, for the single network level, but also for the multilevel analysis and I think for at least exploratory work I think those models would be quite useful.

NC: I think Tracy is right. I mean there are some good models that allow us to be able to look at how there are not simple aggregates, aggregation of individual levels of expertise that then go to the team level, etcetera. I think some of the interesting theoretical challenges there that have not, as far as I know have not been implemented in terms of methodology, are for example things like the so-called Battier Effect. Some of you may have heard of Shane Barrier, who was a very popular college basketball player from Duke, and then he went into the NBA and he was pretty successful, but he was not, he didn't have the statistics that show a lot of success. So things like he didn't have a large number of points, or rebounds or assists and people kept wondering why is he an all-star? Everything thought was an all-star, but he had no stats to prove it. And so the *New York Times* wrote an article about Shane Barrier referring to him as the no-stats all-star. And then And then Daryl Morey, who is the manager of the Houston Rockets, he finally figured out that, he ran the numbers in a distinct way and he managed to end up recruiting Shane to come to Houston, and then of course he moved after that. But the important thing was what he found out about Shane was that Shane had an interesting stat that was not used, and this comes back to how we take multilevel teams, that when Shane was actually on the court his stats were not phenomenal, but when he was on the court everyone else on his team had better stats than when he was not on the court, and furthermore that the opposition had worse stats when he was on the court than when he was the bench. So he referred to that as the Barrier Effect, and I think that we need to be able to see ways in which we could model those kinds of phenomena when we look at trying to take individual level and then apply it to the team level outcome. And I don't think that current network methods have good ways of being able to, or statistical methods more generally have ways of being able to model that kind of an effect on how somebody makes everyone else on the team better even if their own stats are not that good.

AD: That's a very interesting point. It makes me think of something that Peter was mentioning to me when he looked at the basketball data where he found out that a second lag actually would, could be important. So it's not only that it mattered who passed the ball to the star for a successful shot, but also mattered who passed the previous, who made the previous pass, which is quite interesting. It's a way of thinking about that.

NC: Yeah.

SC: That's a very interesting point. Isn't that in the case, that in the study of small-world networks there is such a thing as a hub? So this is a node where it's connected, well connected to a lot of the other nodes in the network and having hubs like that in turn facilitate communication and propagation of signals. So in his case he would be like a hub that's facilitating other's performance. So there may be a different hierarchical structure there, different kinds of network characteristics that would allow us to get at the different kinds of team performance, and context probably is important there. There are some tasks that require a certain kind of team organization compared to other tasks.

NC: Right. I just, I mean it's one of those findings where there's an interesting intuitive finding there, but I don't think that we actually put that into looking at team efficacy and identifying that as an important characteristic of a team player.

AD: Good point. So perhaps the next question? So do we see patterns in the dynamic interactions? I think we addressed that a little bit, but is it anything else you might want to add? No? Okay. And can we improve the model estimation and we heard about the snowballing approach. Are you aware of other new methodologies out there? Can we, you know, in what circumstances can we still use an EM algorithm? What does it mean when we go and we use an MCMC? How do we ensure that our models are identifiable? Any thoughts about that?

TS: So I know that there are faster ways to fit some of these models. I think people use MCMC because it's easy, so when you're developing a new model it's kind of an obvious choice in terms of estimation. I know for some of the conditionally-independent network models that I've been working on that I introduced today there are some EM algorithms that are currently in the works, not by me but some of my colleagues back at Carnegie Mellon, that are trying to speed up estimation by using EM. I also know that there are some variational methods that people are using and I'm not as familiar with those. I know that depending on who you talk to there are advantages and disadvantages since it is an approximation as opposed to finding some type of local or absolute global maximum(?). But I think those methods exist. I think computing continues to get faster, so I think that model estimation will naturally improve. But I think that there are people working on ways to improve estimation. I'm just not one of them.

NC: It's just in the last decade, and I'm showing my age now, but there is sort of a three-way sort of tug of war that's going on here, and not only think of a two-way, but it's a three-way tug of war. One is that you're improving; I mean even within the context of …., which is where I have most familiarity, there clearly have been techniques that have been used to improve estimation and to speed up estimation. That's one. The second is you've got more computational infrastructure so that even if the algorithms are not any more efficient the computational infrastructure is going to help with that. And then the third part of the tug of war is that you're getting increasingly larger datasets to work with. And so there's a race on in some ways between these three that all of these, you know, so you have larger datasets, you have a larger computational infrastructure and then you have some efforts, and I think mostly fairly modest, at least in the area of exponential …. graph models to actually speed up the algorithms. So I am really interested, Tracy, in what you were discussing about what the folks at CMU are doing to speed those things up too.

SC: So it's certainly the case that now hybrid approaches are getting ever more popular. So for instance …. versions of the …. that allows you draw on the strengths of MCMC methods so that you don't have to worry about integration over so many dimensions and while still being able to draw convergence [convergent?] statistics from a frequenter(?) standpoint. And another idea or hybrid approach is when you combine …. MCMC type methods with a filtering, particularly filtering methods. So there are certain …. …. where you do need to get revised parameter estimates, but perhaps not on a real-time basis. So with that you don't mind letting the model, or you may not mind letting the model run for three days so that you can get the parameter estimates. But the part where you are trying to update or get prediction on a real-time basis, you can use some of these filtering methods, fixing(?) …. parameters at the values that you got from the last update so that you can get fast real-time update. But of course that depends on knowing the time scale or how often you need to update your parameters and then it may be a trial-and-error thing that you have to go back and forth to try to see whether a daily update is frequent enough or whether it's a weekly update, and how often and at what time point, what timescales are needed to be measuring the data to be able to do this kind of real-time tracking.

AD: This is also related with the issue of time and scale that we've been talking before, so that's also quite interesting. Well, with this I would like to open the floor for discussions for the audience, so if you have questions please join us. Paul?

R: This is going to be a little bit out of left field. I'm not a member of any of the groups represented here and so that's partly why. And it's a suggestion, and I just want to put an idea in your heads, all of you. I've been sitting here now for a day and a half and it's been fantastic. I don't know how much I've learned, but I've certainly been very intrigued and stimulated by what you folks have had to say, and particularly in the last couple of sessions a though has occurred to me that I want to share with you. When it comes to complicated interlinked networks of the type that we've been talking about this morning legislatures are a wonderful example and it's a new community for you. I mean that would be the political science community, and I don't know how much folks around here have to do with the political science community. Hopefully you do, but I don't see them represented here necessarily. But I was a Congressional Fellow many years ago and so one of my recommendations for you is if you have students or young, new PhDs, the Congressional Fellowship Program is very cool and it would teach them a tremendous amount about the sort of thing that you're talking about.

But just for a moment think about, let's think about the House of Representatives, but it could be any state legislature or the Senate or whatever, 435 members. They are each of them from particular districts, from particular states, from particular regions of the country, so they have geographic connections. So they urban, they are rural. They belong to, as you all know they belong to parties, yes, that's very standard. Everybody knows about that. They also belong to committees and subcommittees, and those are teams and these things cross lines. They also belong to even less well-known things like caucuses. There are informal and sometimes formal associations of members around particular issues, around particular geographic sorts of things, and, of course, around what particular committees they're on. The point about this is that; oh, the other point is that the data is available. These people do everything; everything is in the form of either legislation that passes or legislation that gets noted in some other legislation, then passes or whatever, but it's all public information, so there's a wealth of data and it also goes back a long way. And it would be fantastic to be able to take a look at what goes on today versus what went on when I was a Congressional Fellow and get a sense of the time shifts and what motivates, what represents winning from the point of view of the individual, or of the caucus, or of the Congress as a whole, or the nation or what have you. Anyhow, just think about that and maybe you can get somebody to do something.

AD: Thank you, Paul, especially today, November 4th, so we have some work to do.

NC: I can see that there is actually a considerable amount of work that is looking at Congressional networks, if you may, and teams within it, and in particular I'll point you to the work of my colleague and good friend David Lazer who is at Northeastern University, and he has done extensive work looking at various kinds of teaming, if you may, and networks that engage, that extend within the US House of Representatives and the Senate, and he's also looked at, combined that with using, the use of social media and to what extent that plays a role in the ways in which they project their image back to their constituencies, etcetera. Another piece of work that your comment, Paul, brought to my mind is some work that was done a few years ago by Katie Faust and John Skvoretz. What they did was they compared very different types of networks to see what kinds of similarities there may be in structural signatures and structural patterns within these networks. And they were looking at networks that were as diverse as networks; one of the networks they looked at was co-sponsorship of bills in Congress, so a network tie between two senators, if the two of them co-sponsored a bill. But they looked at other networks. They looked at trade networks across the world where the nodes were countries and which countries traded with one another. They looked at animal networks, primate networks, which primates, which monkeys scratched one another because that represents a hierarchy, a pecking order, which is in fact the term that is used amongst chickens, and then they also looked at the cow-licking network, which cows lick other cows, which also represents a form of hierarchy. And the punchline here is what they found was that the two networks that had the greatest amount of structural similarity was the co-sponsorship network and the cow-licking network.

AD: I like that. I don't know what is their opinion about this, but I like it.

NC: I think the cows must have been offended by that one.

AD: Thank you. Michelle?

R: Hi. I was intrigued by Mengxiao …. work with the dynamics of looking at the task properties in relation to how they affect the teams, and I was actually thinking about this in the previous panel when we had the very different examples of submarine teams working together versus a mentoring situation. And while this is all collaboration, they're very different kinds of collaboration. I was wondering what other work has been done at looking at those task variables and how they affect either the parameters of the collaboration or what we would expect as positive outcomes from that collaboration.

NC: So I mean this goes back to the first panel yesterday and Eduardo's reference about the important of context in understanding these different kinds of things. One specific study that we have been involved in, and actually Mengxiao was also part of that analysis, was we were looking at data from a website that's called nanoHUB. This is a National Science Foundation, NSF-funded project that is essentially, you could think of it as Facebook for nano-scientists. It's actually housed at Purdue University and it's basically a place where people who are interested in nanoscience will log in. They will put software tools, they will put instructional material, course materials, video, datasets, etcetera, and then they will tag it, and rate it, etcetera. And one of the things we looked at there was the assembly of teams that came together to develop software that was being used in the nanoHUB. So there's a software that is, you come together, you develop it, you host it at the nanoHUB and then people use it and it's great for performance. You get outcome measures because you know how many times people use it. There the important task variable that mattered a lot was whether the software was open source or closed source. And we were puzzled because we found that they are some of the most; the people who had the highest expertise were staying away from open-source software as opposed to going to what's closed source. And I had the opportunity since I was working with nanoHUB; this is a very big conference, a big enterprise, and they have an annual conference of about 6,000 people and they call it, because it's nanoHUB they call it HUBUB is the name of their conference, and they invited me to give a keynote and give some of these results and I said, "I don't understand this, but it looks like some of your best programmers are not going towards open source. Rather they're going towards closed-source software." And immediately everyone said, "Well, of course that's the case because the academic incentive system recognizes closed-source software, and so if they are smart people they won't be spending their time doing open source because they don't get as much credit for it." So there are sometimes just incentives existing in the social system that drive you towards one kind of task versus another kind of task.

AD: Tracy, do you have anything? No? All right, thank you. Steve?

R: One thing I noticed after this panel and this morning's panel was the importance of representations, visualizations to convey the complexity of the data, and I'm curious to see the degree to which people in this area of research pay attention to how they develop these visualizations to convey that. And I bring it up in the context of people like in the computer-supported collaborative learning community where when they look at process data, when they look at their chat data they do pay a lot of attention to the visualizations that are created to help understand the data. So this is more a meta-scientific question from the standpoint of how do you use visual representations of this kind of complex data to help you and others comprehend these kinds of findings?

SC: I'll answer that. Visualization is certainly a very, very important component of model building. It's my personal opinion that in constructing any model, before we have a confirmatory model it's important to look at plots of the relationship among the variables and not just in a static sense, looking at how they co-vary together over time and in developing differential equation models, different equation models for instance. There are tools that allow you to visualize smooth …. estimates to get a sense of what are the variables driving the change mechanism and how there may be a moderating effect of some other variables in driving change. So I certainly think that visualization is very important. Also related to that, the modern exploration phase is also very important. So there are certain models where you may want to incorporate measures at the raw item level and try to see how different items are related together and how they may lend themselves to higher-order constructs because then moving on to some other more complex models. So visualization is certainly very important, model explorations and this is not just a one-time thing. Perhaps there is an ongoing process where we need to go back and forth between explorations, confirmatory model and then building or incorporating more exploratory information.

NC: The first word that occurred to me when I heard your question, Steve, was exactly the same issue, that it's really important at the exploratory phase. And it's not just important, but I think it's necessary because very often you can go into very complex analysis and then find that there were some real serious aberrations in the distribution of the data which if you did not look at initially you would have missed it. That said, one of the things that, where we used a lot in Mengxiao's work that I presented was in the area of correspondence analysis. But there again you have to remember that you are reducing the correspondence analysis to the two principal dimensions and you have to make sure that that is capturing enough of what this variability is that you're looking at because it becomes obviously, you know, when you're doing correspondence analysis the map is going to go into N dimensions, but you're at best able to do three dimensions in most cases and in a plot like this two dimensions. So you have to be careful that when you're looking at a two-dimensional correspondence analysis plot, in fact for that matter any of the NDS [MDS?] kinds of plots, that you have to be mindful that you may not be able to see the entire picture when you're looking at that from an exploratory standpoint.

TS: Yeah, I just want to add that especially working with networks, exploration and visualizing the network I think is really important. I also think it can help the substantive researcher figure out exactly which covariates are of interest and which type of model they would like to use. So oftentimes when I'm working with people that are substantive researchers I will talk with them about some of the network plots that I have, and it's mostly exploratory, but just asking questions, you know, we noticed this pattern in the networks, or we're noticing this happening in the networks, or are there any theoretical underpinnings that would be driving this type of structure or driving these types of relationships to happen?

AD: One more question.

R: Well, I just wanted to add to Noshir's point as you were describing how you use visualizations and you can look for anomalies, it reminds me that there's research in sonification where they're using that for example in radiology to pick up things that visually you cannot determine, so that might be something that can additionally be considered.

NC: Good point.

AD? Thank you. Now we are approaching the closing of these events and it is my pleasure to introduce Dr. Scott Paris, the Vice President of Research and Research and Development at Educational Testing Service, and Scott is also one of our strongest supporters in the research on collaboration.

Scott Paris

Thank you very much. I'm just going to say a few summary kinds of remarks and the first one really is to congratulate all of you on your sustained attention. This has really been an incredible amount of information in a day and a half. I know I was struggling between comprehension and cognitive overload multiple times. So thank you for your terrific questions, your participation, and thank you for the speakers who packed an incredible amount of information into these short presentations. The format was terrific and I think we should all give Pat, Mengxiao and Alina a round of applause. Very creative, very effective, so I'm really quite delighted that they did this.

I should mention that about a year, a year and a half ago Pat and Alina and I talked about collaborative problem solving and we all saw it as critical to the future of the things we wanted to do at ETS, and so we decided to make it one of our priorities and we began that more than a year ago. So you can see from the research that Alina and Pat talked about, you can see from the new scientists at ETS that have been hired that they come with incredible and varied backgrounds to bring their expertise to this problem. So we are totally committed at ETS for long-term investment and effort to improve the way we define, measure and operationalize collaborative problem solving. We see it as important for higher education, for K-12, for the workforce. We see it as a fundamental way to reinforce the importance of 21st century skills. So we're delighted that it's part of our agenda.

I was going to ask all the ETS people just to stand up so you can see how many of the ETS scientists are here interested in collaborative problem solving. There's quite a number here. Thank you. I was going to ask somebody to lead us in the ETS fight song. You can hear the collective and synchronized emotional gulp and EEG spike. So there are synchronized activities that aren't related to collaboration, which brings me to a couple of the points I wanted to make today.

It seems to me that one of the themes of the conference was really on definition, and so we heard many speakers that noted the importance of distinguishing collaboration from learning, from problem solving, from teamwork, from simply being together and working together, and the need for a definition permeated a lot of the discussions. In fact Patrick Griffin said, "We always start with definitions of the construct." I thought that was terrific. Eduardo reminded us yesterday to get back to the basics about who, what, why, where, when for the collaborative problem solving. And as an outsider to this research area I was really struck by Steven's clear examples of the micro, meso and macro levels of analysis and collaborative problem solving of multiple players, multiple systems in the way people collaborate. I thought that was terrific. In fact he just mentioned the visualization and I still see the light at the end of the baton and the elbows and the orchestra. That was a great example of coordination among players.

I like the definition of collaboration as a concept. I thought people rallied around the notion of a multidimensional, multilayered approach to collaboration as a concept that subsumes a lot of processes. I think Saad called it an umbrella term that really captured many different processes, and in fact the discussion and the measurement today talked about how many different aspects of networks, of teams of collaboration, of dynamic interactions that we can investigate. I thought that Steven's three-dimensional space of scale, virtuality and interdependence might be a really nice organizing framework for locating different measurement tasks, different paradigms that we use because there's such a diversity of our approaches, submarines to teachers to so many different contexts that we need some kinds of representations of similarity among the tasks, among the processes that we're looking at, so I like that framework a lot. So one thing is then the umbrella conceptual multidimensional nature of the concept.

A second was that people really didn't bite on Patrick's invitation to assess a trait of collaboration, and in fact there was quite a strong discussion of the conditions and the conditionalized values of collaborating. Steven talked about coordination, and a lot of people talked about the dynamics of teamwork and the dynamics of the interaction and not any kind of a constancy in the way people collaborate. And in fact, Noshir's discussion of Mengxiao pointed out the variety of factors that go into the attractiveness of people in groups and the personal biases and selection, "Oh, no, not person A. I'd rather not be on that team,” that in fact we have to look at all the situated and conditionalized variables that affect the formation and the maintenance of these collaborative actions and the teams.

Leslie described these conditionalized things very well with her confluence and her countervailing forces, and I was particularly impressed with the notion of distinguishing team and multi-team kinds of influences. I thought that was a very valuable expanded framework because a lot of us come to collaboration thinking of teams as a small unit and not intersecting with multiple other teams. Although I have to say the question that was raised by Paul about the state legislature reminded me that collaboration isn’t always effective, or successful or accomplished. It can lead to inaction and incompetence at times too, and I thought the cow-licking analogy was very good by proximity.

So one of the things that we have to look at then is the context of collaboration, the outcomes of the collaboration, the dynamics of the collaboration. Noshir described contagion and personal inertia as negative forces and representation, gatekeeping, neutralization as positive team membership. Tracy talked about her network analysis of friendship in group membership. And I think that it’s really important not to just talk about process measures, but to talk about the kinds of variables that influence the membership, the maintenance of the relationships and the evolving nature of those relationships.

So it seems to me that there was one more dimension that I thought was really interesting and that was Vincent when he said, “What is collaboration good for?” And so there’s a functional approach to collaboration that I thought was interesting so that the definition of the construct has to have multiple dimensions. It has to have context as a variable. It has to have conditions for the interactions. It has to look at those changing interactions over time. It needs to look at the functions that the collaboration serves well or poorly. As Lei Liu pointed out productive failure can be one of the consequences or one of the dynamic processes of collaboration. So a very rich, very dynamic kind of an approach to collaboration rather than any kind of a static approach, which makes the second theme even harder and that is measurement.

If we're not talking about a stable trade, if we're not talking about something that's easily defined, and we talk about something that's measured in wildly different groups at levels from EEG patterns of submarine operators to behavioral networks and friendship analyses in classrooms, the levels of analyses and the measures really make comparability of a construct difficult. So breaking down the construct of collaboration, coordination into its components so that we can compare at least across some similar dimensions or processes I think is going to be useful. Eduardo who is a rich source of quotes said, "Measurement is not sexy," but among all the data lovers here I think today we showed that the design and measures of collaboration are indeed sexy and certainly very, very rich and imaginative.

So just in quick review, we had, Vincent pointed out, multiple data sources, tutor logs, iTracking, learning curves, screen captures, art, and Saad added cameras, sensors, gestures, faces, conversations. We had a lot of implicit neurophysiological measures. We had a lot of group dynamics, transactive kinds of measures. Art and Alina talked about multiple interactive environments. And so in fact was it Noshir who talked about the offline world? So it's not a real world and an online world, it's offline and online now. But the environments are so different with human interaction versus simulated environments and interaction, and we talked about lots of different kinds of research across those domains. Trialogues, virtual environments, I think Alina surprised everyone with Holodecks and augmented reality as a new kind of venue to look at this. And then many people reminded us of the particular liabilities in looking at agents versus humans in the interaction. The virtual Carolyn was more interesting than the human Carolyn, and Art talked about people willing to have conversations with a virtual agent that they wouldn't have with a human, so really interesting interpersonal differences between agents in an environment versus humans. Saad and Art talked about using multiple technological tools, Skype, sociometric badges, connects, neuro-measures, chat boxes. Carolyn talked about the discourse analyses and the data mining. Art talked about the butt sensor, which is memorable. So gestures, body position, orientation, all the subtle implicit measures that we can get for sustaining attention, membership and collaboration I think are really inventive.

Then when we started to talk about measurement models yesterday and Alina put up so many so quickly, I couldn't even write them down, but the hidden Markov models and the Bayes Nets that were discussed today by Ron and Peter I think are fascinating analyses of time series data with many mutual dependencies, and how we analyze these very complicated events within these complicated interactions is amazing to me. The last visualizations that were put up with, and I can't even tell you what those patterns meant, but you could see that some patterns were different than other patterns, so it starts to invite questions about how you can plot things to visualize increasing interactions and connectedness over time.

Carolyn described a general framework for transactive knowledge integration with transactivity, engagement, authoritativeness. And again for a relative newcomer like me I found an overarching framework like Steven's and Carolyn's helpful to try and compare different approaches to collaborative problem solving in terms of their relative emphasis on these different dimensions.

The other thing I want to say about context is that I think when we think about scaling up we tend to think of online environments, simulated environments where we can exercise more control, but I worry a little bit that we're losing a lot of the human qualities and the fine-grained distinctions that teachers and students make, that babies and mothers make, that organizational teams make in the workforce about qualities that determine how much they sustain their interest in membership and how hard they try to participate in collective work. And so I think we need to be careful when we extrapolate to environments we control that are simulated that we don't lose too much of some of those special implicit human affective characteristics that make groups either attractive and functional or less attractive and dysfunctional.

So Saad said we need a full spectrum of tools. Vincent said we needed to combine technologies. Art said we had to include multiple measures such as emotions and affect. Carolyn told us about analyzing how people interact differently with agents and people. Will asked a question about reliability across these multiple contexts, which I think is a very thorny problem. And then Noshir talked about the Shane Barrier effect, which I found fascinating. And so if you ever want to get into fantasy sports, I think see Noshir and Peter for some of the statistical techniques for analyzing players. But we clearly need to have a really full toolbox of methods, of measures, and we have to be inventive, and we're probably going to have to stretch traditional psychometric notions like reliability, generalized ability and even the notions of validity because so many of these measures are highly contextualized and bound to situations.

The last theme I wanted to emphasize was the significance of this work. I think the significance was discussed mostly for education. Now Patrick Griffin's work really was the pioneering work for education, and in fact globally when people think about measuring collaborative problem solving they think of the kinds of tasks that he invented. They think about using chat boxes. They think of online environments. I was really struck yesterday when he said imagine how different things would be if he had chosen three people as the unit of interaction instead of two, and it really would have been different, but that's one of those steps that we may take going forward because a lot of education is tutoring, as we saw. Yoav's presentation of tutoring results, a lot of the dyadic interactions, but we need to then look at triadic and tetratic and multiple people involved in these teams.

For the educational significance then I want to also mention that ETS has been a leader with the PISA 2015, and Paul and Eric summarized I think very well this innovative task that they developed given a construct and given the constraints of more than 60 languages, given the constraints of technological delivery. And so to create a way to compare collaborative problem solving across diverse countries and students I think is really ingenious and I'm eager to see the data. As Eric hinted at that maybe the field data suggest that it's not going to correlate .85 and not be just another ersatz reading measure, which is quite exciting, that maybe we're really getting a handle on it. Patrick Griffin described the overarching framework, the processes, the skills, the progressions that they outlined in their method and I think that's a really important framework for the rest of us, that we have to have developmental progressions, or learning progressions or some kinds of novice-to-expert rubrics or stages to talk about the kinds of things that we're measuring so we simply don't have a criterion-referenced or a norm-referenced system for every separate study and measure, but we have a broader conceptual notion of collaboration, or coordination or whatever the construct is and how it develops to expertise.

Let's see, the educational value I think was also discussed in terms of going beyond the traditional measures of reading, writing and reasoning, and people talked a little bit more about the humanistic values, the 21st century values. It reminded me of the holistic education that's now the rage in China and Asia that for those countries as Patrick pointed out that are already near the ceiling on traditional measures of academic expertise and testing proficiency, they're beginning to attend to the other skills that will be differentiators. And that brought up some interesting discussion. Will Singapore turn into test-taking classes for collaborative problem solving? And of course the answer is yes, and of course those classes started years ago. People in Singapore know *kiasu*, the fear of falling behind, and Singapore and Shanghai and Korea are not going to fall behind, so collaborative problem solving will be built into the educational systems of Asian countries. Hopefully the rest of European and American, North American countries will follow it and take the positive advantages of collaborative problem solving for the educational value of helping people learn to be effective in teamwork and group work.

There were some serious questions raised about gender effects. Another question raised, do parents want collaboration or do they only value individual achievement? Another question was raised, will the data be misused and analyzed when it becomes available? I think Alina's answers was of course it will be abused like a lot of large datasets. But one of the positive consequences I think for education in the workforce is the increased attention to the dynamics of these collaborative forces that are so important across education and the workplace, and more attention hopefully will lead to positive consequences and not just invidious comparisons or trivial preparation for the test. If we actually get better at the 21st century skills that Patrick, and Barry(?) and Intel, and Cisco and everybody worked on for the 21st Century Project, if we really get better at those we're going to have significant and positive effects across the globe on education.

So let me just conclude by citing Jay Goodwin's opening. He had a timeline that was optimistic, but in 25 years we might have the technological sophistication to operationalize measures of collaboration and collaborative problem solving. We hope we can do that at ETS. We hope we can produce significant milestones before that, and we are eager to join with you in that progress and hope we can achieve it together. Thank you very much.

Patrick Kyllonen

I just want to cover a few points before we break up here and one is that there have been questions about getting the talks, and as you heard, we are coming out with an edited volume next year sometime. There is going to be a website put up very quickly. The website will have the videos very quickly. We're going to have some version of the presentations relatively quickly, although that has to go through a round of making sure that no one gets caught with some embarrassing pictures or something on their presentations. And I'd like to; and so this will all be up very soon. Email any of us and we'll tell you exactly when and where.

I want to thank ETS. Scott Paris, as you could tell, was one of the great supporters for this work, so he actually was one of the ones who provided funding support for this conference. I also want to thank Army Research Institute, Jay Goodwin, Andy Slaughter is here, and they were also a major contributor to sponsoring this meeting. And then finally I'd like to thank the people who have kept us on time, Liz Coppola. Mary Lucas and Katie Faherty, they're not here, but they did a tremendous amount of work in getting this all together. And then the people who are here, Pia Washington from Adecco, Grace Espeut, Karen Pereira, Nick Sferra, who is back in the back there doing all the videos, thank you very much for all your help.

And then finally I want to thank all of you. Scott's thanked you for putting up with a day and a half of this, but for me it's been a tremendous pleasure and learning experience. And now lunch should be out the door. Is that right, Alina? Yes, we hope. Okay, thank you.

END OF PRESENTATION