These transcriptions may contain errors, especially in spelling of names. These are unfortunate, and we regret that we do not have the resources to fix these errors. Still we believe these transcripts will be valuable to many users.

Socio-Computational Research

Kevin Crowston, Syracuse University

>> Kevin Crowston: Thank you for inviting me. This is my first time at WebShop, I don't have the personal history that some of the previous participants recounted. And I am going to try and talk a little bit about some current work that I'm doing. I'm not going to say anything about NSF because I don't work there. David used to work there. You can ask him, see whether the flashy device actually does all that it's supposed to [indecipherable]. But I was also asked to give a little background about how I got here, so I'm going to start off with a bit of that. So, this is my background. I have an undergraduate degree in applied mathematics with a specialization in computer science from Harvard. This is because, at the time I was a student, Harvard wasn't quite sure whether this computer stuff was really going to catch on and so they didn't actually have a computer science major. And as a result I have a 30 year old knowledge of computer science. The first programming course I took as an undergraduate was P2P11 assembler, and also LISPS because we had to write our own LISP interpreter in P2P11 assembler. And when I was thinking about this talk I realized I actually use many of the things I learned then like, for example, C Shell and BI, more or less have not changed in 30 years. But in fact also it turns out that one's undergraduate degree colors the way you look at the world. So despite everything that's happened since, I still look at the world more or less as a computer scientist does, which is to say gee I wonder how you could program a machine to do that. I did my PhD in Information Technologies at the Sun School of Management, where I switch my focus from technology to the context in which technology is used. Information technologies is what MIT calls information systems. It's not really a discipline, it's more of an application area looking at how people actually use computers and a lot of the work I did there was more informed by organization theory, which is why I tend to think of myself more as an OT person. I then spent five years teaching at the Michigan Business School, which turns out to be a real business school unlike the Sun School, which really is not like more business schools, so there's a lot of focus there on the MBA's, not on the technology. One of the things that was going on at Michigan at the time was that the School of Library Science was reinventing itself as the School of Information, in fact changed its name at the time to School of Information. So I had a sort of sideline view of that going on and one of the interesting things was that they commonly mentioned they were following Syracuse's model in doing that. And it turns out that Syracuse did the same thing except in 1972. So after a while at Michigan I moved to the Syracuse School of Information Studies, which actually calls itself the original information school trademark. And I've been there much longer than I expected to be, so I've actually been there for 16 years now. And one of the advantages of being in the information school is the students that we get actually come to the information school to study information in a variety of different models. We have our [] degree programs but they all focus around information and that's actually very refreshing after teaching in a business school where most of the students were there for like finance or marketing or accounting. I was always amazed at college freshmen who knew they wanted to be accountants, but we had a bunch of those. The other thing that's interesting is that information schools kind of turn the usual disciplinary departments on their head, because instead of having, for example in the sociology department a bunch of sociologists who have different specializations like religion or organizations or whatever, so in that context everybody has the same kind of disciplinary background but they can't really talk to each other because they have very different interests. In contrast, in the information school, we have people with very diverse backgrounds. So we have some economists, sociologists, anthropologists, information systems people, but they all have this common interest and so you can have actually a lot of very interesting collaborations that are interdisciplinary within the same school. So that's actually how much I really value being at the School of Information Studies. And as Mark mentioned, I'm

about to start a different chapter in, let's see, in about two weeks I start at the National Science Foundation as a rotator, and I'm going to be in the Human Center Computer Program which is one that I think a lot of people in this room would probably have some interest in. The way I got there is through my research interests. The basic theme of that actually hasn't changed since I started in graduate school. Actually it did. When I went to MIT I thought I was going study locking mechanisms for distributed databases, but as it turns out the professor I was going to work with went on sabbatical, and so since I needed to work on a research project, I ended up working with a different person, just like a completely different topic. So it is actually a little bit funny how some of those little sort of transitory things actually can kind of end up redirecting your entire life. So instead of doing database locking I ended up doing organization and basically have been looking at new ways of organizing work enabled by information systems for the last however many years that is. During that time, of course, it actually happened. There are in fact new ways of organizing work enabled by information systems so there's always a lot of stuff to study, and I've done that through basically that combination of theory building, empirical studies and some amount of systems development evaluation. And the actual mix between those kind of changed from year to year. And as a result, the methodologies have been pretty diverse. We have a little session about data management. I realize that basically every kind of data that anyone ever thought about trying to manage, and, you know, ranging from the sort of very large scale social media stuff like 500 gigabytes of sort of transactional data about open source software systems development, but also qualitative field notes, interviews, I even have stuff. I went to open source conferences and I collected all the swag that they give out to the developers, so I have a little collection of open source software swag. [Laughing] And in the HCC world, for those of you who have looked at the NSF's website, this is the little picture of the areas that are supported by the Human Centered Computing Program, and they kind of divided off across these different dimensions of the scale of the human system, the scale of the computer system, and the pattered environment that it works in. So most of that stuff you can see I've been up in that upper left hand corner. Those of you who work in the lower right hand corner should not despair. There's actually three programs officers in HCC who, the other two have interests that are more on that side, but this is the side that I'm in. Some of the specific topics I worked with over the years include internet and real estate, genres of digital documents, actually this is kind of truncated a bit. This is the stuff I've worked on with NSF's support. Virtual work, free libre open source software development, data science, system science. In the interest of time I'm actually not going to talk about the first four at all, first five at all, I'm just going to talk about the most recent one which is citizen science. But, if anyone is interested in hearing about genres and digital documents and how that affects work, I'd be delighted to talk about it in more details since that's one of those studies which has been going on forever and which doesn't seem to really be wrapping up at all. I'm going to talk instead about the citizen science stuff. So what is citizen science? Citizen science are scientific projects that involve the general public, and they basically fall into two broad categories, those that collect data, things like eBird, those that analyze data, things like Galaxy Zoo. Actually there are other ones as well. There's a project called Clean up Chesapeake Bay, which might be locally of interest. Which again is citizen science but much more action oriented. In the beginning these were often focused on a formal science education, so getting people involved in the scientific projects is a way to learn about doing science. But increasingly people are looking at them as a way of doing science. To actually get data that is scientifically [indecipherable] or analyze the data. And so we have a couple of projects which are looking at the work practices in these groups and trying to develop technology to support them. So, I'm actually going to skip over this project entirely, even though it's interesting, and get right into this. The thing about the citizen science which is interesting is that it's somewhat similar to other kinds of massively virtual online activities like Wikipedia, but in fact also distinct in certain aspects. So let me just give you an example of one called eBird. eBird is run by the Cornell lab of ornithology. I hadn't realized it but the Cornell lab of ornithology is actually the largest unit at Cornell. It has about 500 employees. It has the world's largest collection of nature sounds among other things. But, one of the projects is about eBird which supports ornithology by collecting data from birders. Actually is anyone in the room a birder? A few people. So, do you use eBird?

>> Yes.

>> Kevin Crowston: So, how do you use eBird?

>> Cataloging what we've seen and where we've seen it. Looking up like if we want to try and find a rare bird, looking up on their maps where other people have reported it during the time of year.

>> Kevin Crowston: Okay, and you also?

>> Super casually. My old place in Santa Cruz where I went to school, I had a bird feeder I made outside my office. And a lot of migratory birds go through there and I would see all these birds stop off at my bird feeder for a couple of weeks and I didn't know what they word. And I thought eBird kind of tried to identify them and then I'd figure out what something is and I started contributing. I only have one location that I reported on, that I would blog what came by my feeder.

>> Kevin Crowston: Right. So a couple of interesting points to take away about that. First of all avid birders are very avid, fanatic. Birding is in fact one of the biggest and fastest growing hobbies in the U.S. You wouldn't think there was a lot of stuff you could sell birders, but, there is. There is stuff going on there, and birders are also list makers. So a very typical thing to do is to go birding and to make a list of everything you've seen. Most birders have life lists. There was in fact a recent hit movie, which I forgot the title of,

>> The Big Year

>> Kevin Crowston: The Big Year. And so that turns out actually to be factually based. The thing that eBird did was to try to collect this data from the birders about observations of different species of birds. And the first version actually did not work all that well, but then they hired some avid birders to redesign it, and what it does it basically takes advantage of what birders want to do anyway which is to keep track of the birds they've seen and to make lists, by making it easy to do that on the computer, and then, so you can see the interface here for example, asking you to submit observations of the birds and then asking them to do a little extra work above and beyond what they would typically do in order to make the data more scientifically useful. So in particular, keeping track of not only what they saw, where they saw it, but how hard they looked. Because it turns out that knowing how hard someone was looking for the birds gives you some information about the observation. So you can, yeah, you can enter the data in a bunch of different ways. Once it's entered you can look at various reports of it. So, they have some which are very topical like when the Gulf spill was going on, keeping track of what birds were likely to be affected by the oil based on observations of the birds. You can also see these range maps so you can get a much finer sense of where the birds actually are. So if you buy a birding book it will have, you know, a map of North America with little shaded areas, but in fact those are not that detailed. And so with the range maps you can see things like, for example, certain birds avoid cities. And so you just sort of see the bird distribution but not in the cities. So that's one example, eBird. So what we want to try to understand was what kinds of conditions can you develop systems like eBird, that can support citizen science? So eBird has been extremely successful. Sort of a particularly good system for one particular kind of science. A lot of insight from the developers who are themselves birders, so we wanted to try to understand a little bit more generally what the conditions were. Develop some models what we could use to describe how this activity went on, and then develop our own systems to see whether or not we understood that properly, so kind of a design science model. So the first thing we did was went off and went to a bunch of websites. We reviewed in fact 27. One of my doctoral students actually did an in depth, her dissertation was an in depth study of three particular projects looking at the entire projects and their organization, not just the technology supported parts, then another part of it was looking at the technology in a lot more detail. There are really literally hundreds of these projects on

pretty much any topic you can imagine. Though a lot of them do have a focus on the ecology. But the second one is kind of an interesting one, Galaxy Zoo. Actually does anyone here do Galaxy Zoo? No.

>> But know about it.

>> Kevin Crowston: Know about it. So Galaxy Zoo is started by some astronomers who had all of these photographs from the Hubel Space telescope of galaxies, and it turns out a very simple question you can ask about a galaxy is what shape is it? And that turns out to do computationally. But it's pretty straight forward for a human to look at the photo and say, oh that's a spiral galaxy, that's an elliptical. So they took a million or so images from the Hubel, they put them up on the website, they figured if they could get a couple hundred people to work on them, that they would be able to get through that data in a couple of years. And it got mentioned on the BBC Science program and they immediately had tens of thousands and then hundreds of thousands of volunteers and they actually blew through all the data within months instead of years. In fact, they were in this awkward position of having a project that had run out of data and that they were concerned about wasting people's time having them reanalyze data that had already been looked at. But it turns out that the next generation of space telescopes will generate as much data as the Hubel did in basically every evening and so that particular problem of not having enough data has been solved. Anyway, one of the things that we found in looking at these is that they had a bunch of different features for collecting data and so on, but very few of them made use of games and contests and things like that to try to motivate participation. So, we decided that we wanted to look at the possibilities of increasing motivation by using game like features. So, lots of projects rely, for example as eBird does, on inherent interests. Galaxy has turned out to be inherently interesting. A lot of the other projects [indecipherable] like they were languishing a little bit because they're about [indecipherable] or mold, and getting people interested in those seemed to be a bit harder. In fact, ours is currently about moths. Are there any moth watchers here? No. Okay. So in designing this we took a design science approach and we have this sort of overlap of things that we want to know about in terms of research. The context and then things we could actually build. And so we end up trying to center our study in the middle of this intersection of things we know about, the context and the build. In terms of what we wanted to know about, basically the research portions, what motivate people to participate in these, how can we increase that as, you know, people have pointed out there are lots of possibilities, but we're hoping that we could tap into entertainment and games. Because we have two birds in the room. How many people play online games? Games on the computer at all? So quite a lot more than two. It turns out that if you count up the number of hours people spend on things like Angry Birds, you have this huge cognitive surplus, which is currently going into shooting pigs. And so the question was if you could somehow tap into whatever motivates that, and get people interested in doing something which would advance science at the same time as it is entertaining, that potentially you could get a lot of cognitive effort applied to these. But then the flipside is if we can design these games so that people will actually be engaged and motivate, will that in fact lead to useful data. Will the data quality be high enough to be worthwhile? Because you could imagine that if it's too game-like people will cheat and they will basically do whatever they can in order to get the score or whatever the prize is, so we wanted to try to understand this intersection of motivating, interesting, fun, but at the same time doing something useful for science. The particular context we looked at, was taxonomic classification. So, very much akin to Galaxy Zoo. Galaxy Zoo is morphological classification, we wanted to do taxonomic classification, and part of the reason there is that naturalists and enthusiasts have thousands of photos of specimens, and in fact a very large number of citizen science projects are actually collecting pictures. So you can take a picture of whatever it is that you've observed, turns out to be really hard to take pictures of birds, but it's easy to take pictures of trees, so there's lots of pictures of trees or various other things, but these are not very useful without taxonomic classification. If you don't know what the picture is of, then it doesn't do you very much good. And the number of pictures, especially in a citizen science context, quickly become much larger that professionals can deal with. So that's basically the Hubel Space telescope problem, but for other kinds of things. And as I said, people are starting to collect them but don't always

seem to know what to do with them. So here's just an example of a small collection of moths that we actually got from a guy named John Pickering. Actually set up by our contact that was mediated by people from University of Maryland. And as you can see there's lots and lots of pictures of people that submitted. Some of them, like this one here for example, there may be moths in that photo but it's not actually possible to analyze them. Some of them actually probably aren't moths like the ladybug, but, or the frog, but there's lots and lots of pictures of moths, but in order for this collection to be of any use to anyone, we need to know what kind of moth it is. And it turns out that there is an approach that you can teach non-experts to use in order to classify the moths. So most of you need to be an expert but non-experts can do this by looking at the particular specimen characteristic by characteristic, attribute by attribute. So there's these different characteristics like the wing shape, the color, the pattern, each of these has several possible values and if you can, and mostly those are simple enough that someone can do them without a lot of training and once you've identified the characters and states, if you have enough of those, you can actually pin it down to a particular family or set of possible species. But not all of them are easy. Some of them need fairly specific knowledge. Like one of them is called orbicular spot, that turns out to be useful for classifying moths, but unless you are trained, somewhat trained, wouldn't necessarily know what that was. And this context also differs from some other context in which people views purposeful game. So for example there's a game called the image game which I think was originally by Luis van Ahn and which is now Google, where you just get people to quickly apply as many labels as possible to a photo in order to have tags on the photo for searching. In that case whatever a person decides to apply as a tag, it's probably a reasonable tag. I, this case, there actually is a ground truth. There is a species that the moth has or we, well, without getting too metaphysical, we choose to believe that the moth actually does have a species, that there is a correct answer for the character state combinations. People opinions about those are not especially useful if they have the wrong answer, hence the question about the data quality and also we have a lot of things which are not all that attractive. There is a theory that the reason that people like Galaxy Zoo is because galaxies are cool and a lot of people like looking at pictures of stars. People find birds inherently interesting and they spend a lot of time and effort looking at birds. But moths, you know, maybe not so much interest. So basically, as I said, you know, experts probably do it as part of the science they're doing, they're enthusiasts, you know, really like this sort of stuff, they're willing to do it. We're trying to go after this much larger group of gamers. And the way we did it was by basically designing a set of games. And we came up actually with a variety of systems, three different systems, that try to aim at different points on the spectrum, and part of the question is whether it makes a difference going after something which is more game-like or less game-like. So the least game-like of these actually was designed for experts. It's the image I basically showed you. You have a bunch of pictures of moths and this is something that an expert could actually use to develop one of these classification systems. So it basically says here's a bunch of moths, which ones go together? What's the character that they have? What's the attribute that's similar in all of those? So, putting together all of the moths that have the same shape and giving a label to the shape. So we're actually envisioning this as something someone could use to basically populate the game, so getting an expert to sit down, classify some of the moths, tell us what the characters and the states are and then we could use that in other parts of the system. But, not something that's much fun right? It's not something that you just do to pass the time. So we have a second system which is actually more [indecipherable] in that way. And this is set up basically as a contest, so we have a bunch of moth pictures and in this case you can see it's asking about the shape at rest and you can just drag the moths onto the shape that they are. And the trick is that one or two of the moths have already been classified, so we know the right answer for that. So we can basically give you a score. We can tell you how well you did. And the assumption is that you know if you do well on the one that we know the answer to then, you know, your data might be good for the others. Yeah, so the shape at rest, four wing distinctive color. We know what color is the moth you can drag the moths that are the same color, and then at the end it tells you yes, you have found one that you classified correctly, and now you can actually take the final step of saying which one does it most look like and what is the species of that one, and then get a score. So in this case you can see that on the first moth I got a perfect score, 80 out of 80 because I got all of the characters right, but on the second moth I got the four wing distinctive pattern

wrong, I said speckled and it was banded. So you get a little bit of feedback, you get a total score, and then because I got the one correct I've collected that one, it goes into my little collection. We haven't implemented this part yet but we're thinking about having things like badges for getting a particular high score, having the ability to compete against your friends, for example, you could have this tied in so that you can see what your friends got and whether or not you're getting a better score than them, things like that. So there's a lot of things we can try on top of this basic game. And it is actually, you know, it's a way of passing the time. It's probably as not much fun as Angry Birds, but it's probably, you know, it's getting into that ballpark anyway. And in fact we'll know a little bit more because we're about to go live with that one if anyone's interested. The website is citizensort.org. You could actually play this one today. It doesn't make you get an account so you don't have to give away any personal information. So that was the sort of mildly game [indecipherable] and actually we're also working on a version that works on the phone so you can play it on your phone if you're tweeting to and from work as long as you have an internet connection. So that was the first of them. But we also wanted to see whether we could attract much more hardcore gamers, so we developed a really game-like game, which I have to switch to show you. Hold a sec.

[Silence]

This is a game we call Forgotten Island, it's a citizen science adventure. And it actually has music, so let me plug in my. So this is set on an island a forgotten island and when you wake up you play a character named Dr. Science which comes in male and female version. And we have, you know, the full kind of, actually, it's basically a flash game, so we have some nice backgrounds you can explore. There are various you have to do. One of the biggest puzzles is just figuring out who you are, because this is your office although you don't recognize it when you're playing the game, and there's a trading element based into it. And what you're trading for is the classifications for stuff that you need in order to finish the game. And actually if you do finish the game, there's actually a little garden that you can grow, so you can keep trade on classifications for garden stuff. It actually has robots, a good robot and an evil robot. Don't tell Senator Coburn about the robots, thank you. And so the bad robot basically gives you your task which is to find these photographs and classify them in order to get the resources you need to finish using the atomic classifier. And the atomic classifier looks exactly like Happy Moths. It's the same task [indecipherable]. It's a picture of a moth and the states and characters. And the way the game economy is set up you basically have to play the equivalent of actually this has probably changed since the game is constantly evolving, but you have to play the equivalent of about 18 rounds of Happy Moth in order to get the stuff you need in order to complete the fantasy game. So the basic hypothesis is Happy Moth, you know, kind of fun, sort of interesting. Let me stop the music since it's a little interrupting. All right, so the basic question is in order to finish a game of Forgotten Island, you have to play the equivalent of lots and lots and lots of rounds of Happy Match. And so the question is will the motivation of the game attract people who had never paid attention to moths at all, who would certainly not play 18 or 20 rounds of Happy Match, but who might be willing to do that in order to see what happens in the fantasy name. So does that tradeoff between the motivation of the game and the additional difficulty in creating the game, is that a good tradeoff. By the way, this is currently about moths but it could be about any photo collection and we're actually working with people from the Smithsonian who have collection of sharks and I forget what the third one is, but there are other kinds of species that are going to be introduced into the middle of this in a bit. And I should also point out that the system was developed by a fairly large cast of masters and undergraduate students. One of the nice things about working at Syracuse was that we actually have a visual and performing arts school that has talented artists that we were able to get those people to help do the backgrounds and animation. Yes, and also aid from people at University of Maryland include Jenny Preece, who's here, and from the Smithsonian and Discover Life. So that's actually been a fairly large collaboration. Just to continue on the theme of citizen science for a little bit, we're also working with those [indecipherable] people, and this actually has been funded, [indecipherable] runs a set of citizen science projects which actually have the same kind of classification at the core, doing things like looking for planets in data about the brightness of suns but also

some nature ones like whale communication. And as I said, this is the Galaxy Zoo project. So what we're going to be doing with them is trying to develop ways to increase motivation but also learning, so learning is going to be a new thing in this particular upcoming project. Learning, in a couple of senses. One is just how do people learn how to do the classification better, so for example with the moth one, the first time you do it, you know, if you don't get a very good score because you're still learning what it means for a moth to be banded or whatever, but over time your performance does improve and so the question is whether we can make that happen faster, for example, by having better exemplars, or giving people little tutorials or what have you. The other is whether you can be more efficient with peoples' time. So, for example as I mentioned, the Galaxy Zoo project originally had the situation where they sort of ran out of data and so they were having people classify things more often than was really necessary. But that's actually a dynamic problem, trying to figure out when getting another classification would be useful. If something is, you know, basically everybody agrees on what the classification is maybe you don't need another classification. If everybody disagrees, maybe it doesn't do you any good to get more classifications because it just such a bad image that there's not going to be any consensus. So trying to work out exactly when you would want another classification is a bit tricky, and then also trying to work out maybe if someone is a novice, you want to give them really easy ones, both because the result will be more useful and also because it will help them to learn. Maybe if somebody's an expert you want to give them the really hard ones to take advantage of what they know. But that actually trades off with motivation. Because if all you show people are really, really hard ones, maybe it gets really tedious and boring and so it destroys the reason people are taking part. So it's a lot of things that are going to get traded off. So basically the idea is to develop a few based on theories about learning and performance. Come up with some suggested interventions into the systems that they are developing and then experimentally test those out, because one of the nice things on the web you can do is AB experiments where you can actually send a random sample of people to a new interface and observe their performance over time and see which one works better. So I think that's what I just said. All right. So things like test and feedback, you know, just telling someone that they got it right, improved performance to telling someone they got it wrong, just increase motivation, so those are things we'll actually be able to test. Anyway so mostly what I've been talking about is stuff that I have been doing. There's some things that I'm entrusted in continuing to pursue which I'm going to use just to wrap up. One is basically, as I said, since I'm interested in how wide spread use of technology changes new forms of organizing, I think citizen sciences is an interesting venue but it's still at a very early stage in terms of the research being mostly phenomenological. In fact a lot of the stuff that we've been talking about through research is very phenomenological. There's, aren't very good theories, there are theories but they're not widely adopted anyway. How's that? And certainly there's a lot that you can see where people are struggling to figure out how to measure things, you know, what is it you can actually measure. In our case, you know, can we measure motivation? You know, there are ways that you can do that. If you have someone in a lab can you do it at a distance? When they're playing a game, maybe a little less clear. I think the previous topic was a nice little illustration of that. We have an enormous amount of Twitter data, what does it mean? So we're still struggling with some of those basic things. So in that respect, I think the challenge is developing a science of citizen science if you will. So building on the research that people are doing on other kinds of communities, contributing to the underlying theories about motivation and learning and, for example, to inform understandings of learning and how people actually learn new sorts of contexts. And as I said, it's not that no one has ever thought about this because obviously there's hundreds of years' worth of scholarship and probably 50, 60 years of relevant social psychology research, but trying to see how that actually connects into these new settings is sometimes a challenge. I'm going to skip that one. I think more generally there's this shift going on which, at least information systems people are a little bit slow to grasp, so historically information systems viewed itself as a field that studies technology use in organizations, and the reason was simple. Because organizations were the only ones who could afford the technology. But now, individuals have way better technologies than most companies have. You know, people own laptops that are as powerful as anything that they have at work. They have cell phones which are in fact better connected than the system they use at work. People have referred to this in a variety of

ways, domestication of technology or digital natives. A lot of companies are struggling with BYOD, which stands for Bring Your Own Device, right. So, they ask questions of the IT department like, if I can look at, fill in the blank, on my iPad, why can't I look at the company reports on my iPad? And there are a lot of reasons for that, some good, some not so good. But, I think the theorizing we've been doing in the past about how people use technology has been colored by that organizational focus and really trying to think about what it means when the technology is widespread, it's distributed, it's owned by the individuals as opposed to be owned by the companies that employ them, you know, I think that's going to be an interesting challenge going forward. And so with that, maybe I'll stop and see if there are any questions. Or perhaps since we're supposed to have lunch. [Applause]

>> Kevin, I'm looking for Forgotten Island but I'm finding your papers, but I can't find, can we look at that?

>> Kevin Crowston: That one is, that one is going through a bit of beta amongst the research team. We are going to have a public release of that one September 21st, I think it is. So Happy Match is sort of in a limited open distribution, but Forgotten Island has some usability problems. I mean, the game is basically there but there are some things that are not quite right about it, so, so that one will be another month. Paul.

>> Paul: I'm curious whether you'll make use of any formal modeling and optimization to the assignment [low audio].

>> Kevin Crowston: Yes. That's the plan anyway. That project is a joint project between us and Adler Planetarium and the people at Adler are the ones who are developing the system and their part of it is exactly what you said, they want to do some formal modeling, they want to use some machine learning techniques to try to say, given a particular image, given a set of users who are available at the moment what's the optimal image to show to a particular user. And then our side of it is much more the social science side of what are motivating factors, what are learning factors that ought to be taken into account, so that it's not just, you know, here's a particular image that has an information deficit, here's a person with a known ability and making a match just on those grounds. So yes.

>> I would hope you would program [inaudible] motivation are into their optimization problem rather than doing it as a separate.

>> Kevin Crowston: Right, that's the intent. But it does say something about the state of operationalization of theories, that we can't just say here's the formula, stick that into your optimization and you'll know what to do. So that's actually a challenge for the development of the project going forward is so, you know, at the end of the day they want to be able to run some equation over their images and on the people who are online and we need to provide them with something they can actually put down into the code and use as part of the optimization.

>> So following on from that, which theories are you particularly looking at? So I think this is a big problem, identifying the appropriate theories from the psychology [low audio].

>> Kevin Crowston: Well, so a student and I have a paper where we try to synthesize, actually something that looks very much like your reader to leader model. We only have three phases but it basically tried to describe the same kind of phenomena, and that is centered around viewing the contributions to systems like this as a kind of helping behavior. And the stages in helping behavior are recognizing a need, perceiving that you have the ability to meet the need, perceiving there's some kind of social force that suggests that you ought to meet the need. However, whether that is actually specific enough to be able to say this is the image that will be most motivating that this user, is not actually all that clear. And so, and it also doesn't incorporate features that the Galaxy Zoo people believe are important, like they make the claim that if you

show somebody a really beautiful galaxy as one of the first three, they're much more likely to persist, suggesting that the reason that people do it is because they like looking at pictures of beautiful galaxies. That one turns out easy to test experimentally, at least, but it doesn't really fit in conceptually with the [indecipherable] which I just outlined, so actually I guess I'm throwing up my hands and I'm saying I don't know. I mean that's why they gave us the money.

>> If I can make a suggestion.

>> Kevin Crowston: Okay.

>> You know something about beautiful images could probably go in there, but how about an optical challenge, so you may be able to make a prediction of the probability that this person would get the question right, [inaudible], should go off as they do more, and it could be that you could actually comparatively test what's the probability that they'll come back given you know how they've done in the last five. I mean it could be getting 80 percent right could be more motivating than 90 percent or 70 percent.

>> Kevin Crowston: Yeah, well there's actually another piece of that which is built into what you're saying which is interesting, which is does it make sense to tell people how well they are doing. And then there's also feasibility, given that people are analyzing images that have never been seen before, how do you even tell them how well they're doing. So feedback is definitely one of the things we want to look at. And as I said the challenge is not only what kind of feedback and how often but just the logistics of how do you create feedback for novel tasks when people are doing something that, by definition, has not been done. But what we were thinking we might experiment with is agreement with consensus. So in other words on the last one, six other people thought the same that you did. But it's possible that that could be motivating. Oh good, you know, I'm on the right track. It could be de-motivating, gee if they already had six people look at it then they didn't really need me. So, that could, I mean it could really damage this feeling of need, you know, they don't really need me, they've got a lot of people, everything I've looked at six other people have already looked at. So.

>> Well, at that point you would throw in like say oh this one's different, or this one's a special challenge or special probe and this is the one that six other people have looked at. But you don't necessarily tell them that until after the complete it.

>> Kevin Crowston: Yes. Well there's lots and lots of design variables and so the way that project is set up is the first year, actually the first year is much more exploratory. We're going to spend it looking at their past data, because they basically have every click that anyone has done on their systems, so we can actually look to see whether we can identify patterns in people who persisted and those who didn't. The other is doing interviews with people who are involved in the project to find out what they claim as their motivations to understand their actual use behaviors a bit more and also to try to look at the learning. What kind of learning they do, because there's a very low level of learning which is the just being able to classify images faster. Some of them, they have different kinds of images like there's one about reading Greek, so building up your recognition of what a hand written Greek alpha on [indecipherable] looks like. But then there's higher level learning about how science works or the nature of the specific science, you know, about whale songs or about planets around stars and it would be interesting to know whether people are learning at that higher level, and so we want to look at that as well. So again, trying to get a sense from the actual participants what they think they learn and how they learn it so. But basically it's a big hairy, yeah.

>> When you use these games, what was your, what is [inaudible]. So, [inaudible].

>> Kevin Crowston: Well, the games, the moth classification games are not ones at the moment that we're thinking about very dynamically, but the questions that you're asking are exactly the heart of the next project which is the Galaxy Zoo one. And the hope is actually that whatever we learn in that context we could fold into these as well, but the moth one actually ends at the end of this coming year, and so exactly what its feature is, is something that we have to figure out during the year. But exactly as you said, trying to make a game which is more adaptive to the user so that as you play, it doesn't get boring too quickly, you know, that it keeps, for example, increasing the level of challenge, the way that Paul was suggesting or it shows you different kinds of things. Those sorts of interventions are exactly what we want to try and explore in [indecipherable] project.

>> I'm sorry. So do you want them to spend more time or do you want them to get something out of it? That's really what I was most interested in. Because [inaudible] people from using because there are, [inaudible] there are limits to what people spend on games.

>> Kevin Crowston: Well, from the point of view of the citizen science project, there's a large pile of photos that need to be classified, and if it turns out that a small number of people, I mean you would expect a skewed distribution of effort because basically actually if someone wants an interesting theoretical challenge, explain why we see the exact same distribution of effort in Wikipedia, open source, game playing, I mean they all have exactly the same curve. And the thing which I find interesting about that is that if you look at an individual open source project that has say 20 or 30 developers, it means the person who contributed the most has contributed a few hundred or a few thousand patches, it turns out that you see the same curve for Wikipedia, which means the person who has contributed the most to Wikipedia has contributed something like 150,000 edits, or some insane number like that. But it's, you know, it's the same curve. But anyway, you should expect to see that. So then you're question is am I responsible for the mental health of the person in position one and two and I don't have an answer to that. I mean, they're going to go...I think it come back to what I was saying. People are going to spend their time doing something and if it's something that has societal value or at least value for the person who owns the moth collection, that might better than having them spend the time playing Halo or watching T.V. or counting trains or whatever else they might have found to occupy their times instead. But that, right. It's an interesting ethical question which, you know, I don't think if you ask the people at Farmville, they would have much better of an answer. If fact they would probably say we call those people profit centers. [Laughing] Each one is assigned an individual representative. [Laughing]. But you know, liquor stores make most of their money off of the one percent who come in once a week.

>> I was just going to say in sort of response to this, thinking about sort of drive [low audio] being played, are you familiar with [inaudible]. He wrote a book called "How to do things with video games", and he created this Facebook game called [inaudible], and his idea is that

>> Kevin Crowston: It was a parody.

>> and people will do it because they, because, right? It was out for a couple of years

>> He was trying to make it more and more annoying to play.

>> How tedious the game was to try to see if people would stop playing

>> And they weren't.

>> They just continued.

>> So I think like the mental, our, I don't know if we should be concerned with the mental well-being of people who just generally continually click on something.

>> Kevin Crowston: So that, that particular sentence is one that is not going to appear in IRB application. [Laughing]

>> Yeah, I would edit that one out, so.

>> A serious answer to this question that we might be hard wired to go ahead and [inaudible]. [Low audio]. All right. I'm between you and lunch, so thank you. [Applause]