

## Thesis Proposal: Next Steps

Nitin Sawhney, March 1, 2002

What should the thesis research focus on?

In my proposal I began with the premise that collaboration is necessary for problem solving in critical domains where expertise may be distributed. I presumed that the real problem inhibiting collaboration was the lack of awareness and sharing of relevant knowledge among stakeholders of problems and domain experts who could help develop solutions.

Based on feedback, this approach could be considered naive in some of its basic assumptions:

- *That stakeholders cannot solve their own problems, and external expertise and technology is necessary.* This may be true in some specific instances, but many would claim that meaningful and well-established (indigenously developed) solutions to problems usually exist within the community, but may simply have been inaccessible for a variety of social, economic or political reasons.
- *That "Collaboration" is a necessary approach for solving such problems; why would the communities, innovative individuals or organizations (non-profit or entrepreneurial) not be able to tackle such issues on their own. What incentives would motivate them to initiate collaboration? After all collaboration requires establishing trust and intellectual property or economic arrangements that benefit all parties. Under what conditions have groups undertaken cooperative initiatives? What does the literature reveal?*
- *That a distributed online system will allow all parties to share and establish cooperative arrangements easily.* Even if all parties have readily available network access (not always true), what would incentivize them to work with unknown participants? Why would they share their expertise or even problem specs with others?
- *That an Open Source approach will allow all parties to contribute and work together.* Anil Gupta pointed out many situations that can kill the open source arguments.

### Potential Research Directions:

Without getting into any detail at this point, I just quickly list below the 3-4 research directions one can pursue:

1. Case studies on existing examples of cooperative design among distributed individuals or organizations. Study how/why open source software development worked (many case studies documented), and how the conditions differ in engineering design. More specifically examine examples of hardware development initiatives – where open source could be implied – sporting equipment (Windsurfing/Skateboards) from the Sloan 2000 study or the Simputer project (I have ready access to the site i.e. the founders and developers in Bangalore). Consider the conflicts arising and how they dealt with intellectual property issues.

*[ This by itself is probably too much of a Sloan/DUSP study? What makes it suitable for a Media Lab thesis? Perhaps if its results are integrated into system design criteria that justify the need and requirements for online cooperative systems for engineering design. ]*

2. Examine evidence of cooperative innovation among rural innovators in the Honey Bee Network. I'm not sure there are any examples or sufficient data on them. As they usually capture innovators who do not take outside help to come up with their innovations – Honey Bee promotes the individual innovator. In any case this would mean spending more time in Ahmedabad, India (which may be more appropriate in the summer).
3. Study the nature of cooperative projects conducted in the DtM Studio Courses at various sites. Don't know to what extent they will use online site. The Bangalore group does not seem to be leaning towards using ThinkCycle as they don't want to put their projects in

the public domain yet; they will initially use HoneyBee and later provide some data on ThinkCycle. The projects at MIT and other schools may use the site sparingly but not to an extent that would allow me to examine complex factors relating to participation from external groups and Intellectual property (because as Bish pointed out, the stakes in class projects are not too high). A potential lack of “critical mass” of users/projects on the site concerns me the most. So what would be the best-case scenario one can study here? Certainly I could try to coerce faculty in 1 or 2 schools to encourage the students to use the site in specific ways as part of our study. But this would only be worthwhile if I have worked out very specific outcomes/measures to study (as Mark requested).

4. In addition to the current ThinkCycle platform, develop more design tools for collaborative engineering design – (is this more of a Media Lab thesis?) but there may not be much chance for social evaluation here. Potential tools that would be useful include – online Hardware CVS system, temporal design visualization, or intelligent search and matching of problems/concepts/interests. Of these CVS is probably most crucial (version control and management would allow large scale and serious collaborative projects to be developed), and the visualization tends to excite most folks (but is very difficult to implement; could build a proof-of-concept prototype only). The non-ML committee members may not find this in itself a relevant thesis from their perspective?

Finally, one could consider some combination of 2-3 aspects above, which complement data and hypothesis from the others, if no one is sufficient to yield interesting results by itself. For examples some combination of 1, 3 and 4 could make for 3 separate chapters of the thesis with another chapter that bridges their results together. For part 3, I can select 2-3 schools – MIT and Bangalore and/or Portugal (since they seem the most serious) and look at differences in their approach and process/outcomes of cooperative projects.

*What I can do in the next 2-3 weeks?*

- Develop study outcomes and measures – need to find related studies for prior approaches. Must be able to setup study and engage the participating faculty in reviewing, adapting and implementing the model for the study (2-3 weeks).
- Finish making current ThinkCycle system changes for a stable and usable version 1.0 release (2 weeks). The system is nearly ready now; I just need to add minor features, and fix a few small bugs and optimization issues.
  - Hold-off on any major software development initiatives with CVS or visualization till later in the semester/summer as it won't be ready for the current course students to use anyway.
- Based on feedback from committee, finalize thesis proposal and submit to DCGS (2-3 weeks).

*Timeline:* If things proceed accordingly, then I can spend April and May heavily monitoring system usage, conducting surveys, observing design teams, and interviews (including some site visits). In the summer I can compile and write-up results and perhaps build additional software or make design changes based on user feedback. I could also study other open source cases (for chapter 1). Presumably I would have enough to put together a more coherent picture and could plan to defend sometime in fall this year.

**Note:** I have not attempted to answer questions posed by the committee here; these are difficult questions and clearly should guide my thinking and assumptions in the study. Hopefully some answers will emerge as I delve into more prior literature in the next few weeks and begin the study component. I recognize that I still need to frame the thesis proposal appropriately and list my key hypothesis, but this is where I need more input from the committee.

## Dialogue with Thesis Committee

Nitin Sawhney, Feb 21, 2002

I'm including here a summary of feedback I received from all committee members, based on what I recall most clearly from my scribbled notes. This is an evolving document; I will continue to add comments as I have further dialogue with the committee.

### **Sandy Pentland, MIT Media Lab, Oct. 29 and Nov. 21, 2001**

Sandy Pentland suggested the notion of **design syntax vs. semantics**. I inferred from the discussion that his notion of "design syntax" might be a way to categorize social behavior and related design activities, independent of the actual content contributed, which may be considered the "design semantics". Though both concepts are clearly inter-related, this distinction may be helpful in analyzing the nature of design projects and the support that collaborative tools can provide in the process. The use of syntax vs. semantics in human language provides a means of understanding the underlying structure and how it influences the generation of the semantic content. *How should one seek to examine the design syntax in a collaborative design project? What parameters to consider and how should they be measured? What would it reveal about the structure of the design activity, and why is that relevant for software design tools?*

*I then speculated on a few potential aspects of design syntax that one might examine based on the (sparse) data that we expect to gather from several design projects on the collaborative tool:*

- *Identifying key stages in collaborative design process i.e. negotiation of requirements, searching related work, exploration of design space, convergence on design path etc.*
- *Identifying the prominent design contributions that influence the overall design trajectory by examining its role in the design structure and the number of threads referring to it.*
- *Identifying the primary design threads and contributors that most influence the final design of the project.*

Sandy suggested that I do a statistical analysis of parameters such as type and time of contributions, thread length, thread structure, use of private threads etc, that may be used to understand the variances and similarities in the design syntax among different projects. Sandy suggested that I examine measures of randomness and entropy or coherence in email conversations, to imply something about the design process. Could one adaptively adjust the user interface; how would that modify the ongoing discussions? – *However I just don't think there is enough data to examine. Such a framework may indicate means for enhancing collaborative tools to better facilitate the design process by providing informative feedback to users or generating suitable visualizations of the design space. However, statistical analysis of such a domain requires a significant amount of data from hundreds of postings over a long duration by several design teams to reveal meaningful correlations.*

On a second meeting, Sandy asked me to narrow down the key hypothesis/goals of the thesis proposal. The best I could come up with was – *the following aspects are necessary to support Open Collaborative Design in critical problem domains:*

- I. Lowering barriers to sharing and contributing design knowledge during the design process, i.e. developing online tools that support informal and lightweight interfaces to sustain collaborative design activity.*
- II. Capturing and representing the evolving design process and informal design rationale from participants in an unintrusive manner.*
- III. Supporting mechanisms for engaging an extended network of diverse participants in different institutional settings, including social incentives, roles and policies.*

This was listed in my proposal ... but apparently it has far too many assumptions and remains an unfocused or unverifiable set of hypothesis. So it has to be rethought!

***Mitch Resnick, MIT Media Lab, Dec 14, 2001***

Mitch referred to Amy Bruckman's thesis; how community strengthens design and vice versa. She did case studies examining interactions among two or more participants over time.

Mitch suggested that one could develop a taxonomy of various types of collaborations observed among participants. There may be 2 or 3 models, based on patterns of interaction? One aspect to examine may be how design criteria for collaborative project got defined.

***Mark Ackerman, U. Michigan, Dec 14, 2001***

Mark felt that I had merely presented a "pre-proposal" for the thesis so far. It needed to be further developed into a real proposal. Mark suggested I put together one coherent page on a detailed hypothesis, methods and approach, including data I hope to gather and expected outcomes.

Mark asked me to consider two key questions:

**Q1.** *What burning thing do you hope to answer/prove from your work on this?*

**Nitin:** We've been hyping this notion of "open source collaboration" for engineering design. A lot of folks seem to be excited by it. But I'd like to find out from actual design projects, what conditions actually support such open source problem solving and design? In many cases, I expect it just won't be socially sustained with distributed participants.

\*\* I want to prove that its possible to have a Linux-style distributed developer community around critical engineering domains like 'low cost water filters', if certain conditions are in place. What are those conditions???

It may be that only certain problem domains are better suited for it, due to the nature of the impact or intellectual outcomes (which are not suited to patent) or that some social/tech affordances of the collaborative system itself will allow outsiders to contribute to the online Knowledgebase and allow cross-institutional collaborations to thrive.

I think we have the opportunity to examine some of these factors in this particular initiative (especially if many design schools actually use it next term) ... but I guess I need to be very clear what I'm looking for ... and how it would emerge in the field-work?

**Mark:** To answer (1), you need a successful adoption and use. That's a lot to do. You will also have to measure (a) success (performance, user satisfaction, adoption of technologies, ...) and (b) the relevant predictive variables (amount of use, social cohesion, similarity of users, number of affective postings versus instrumental postings, ...). You don't have to do this - you could do a straight ethnographic-style study, but you're always better off getting some numbers. Your adviser may demand this. One can imagine lots of outcome measures, and lots of predictive measures, qualitative or quantitative. What are you going to use? What does the literature use? What models of outcome and causation does the literature use?

**Q2.** *In the academic world (not consultants, not World Bank, not corporations), who do you want to hang out with?*

**Nitin:** A. Sociologists doing computer science (and vice versa) ... folks that I currently tend to identify with Intellectually.

B. Faculty/Researchers working with innovative approaches to teaching/collaborating on design for appropriate technologies – Vijay Chandru, Poonam Kasturi, Anil Gupta ... many of these folks came to the workshop I organized this summer. A wonderful peer community.

**Mark:** To hang out with (2), you need to come up with research questions that are your invitation into the (2) groups? I think the first group you mention are CSCW people? The second group is technology transfer people? If so, now you have the relevant literatures. Your pre-proposal had some CSCW literature. What specific questions would a CSCW person want answered about adoption and use? Would this differ from a tech transfer person? What questions (and models) would they use? (I use the model on purpose - I think the tech transfer people are much more empiricist.)

Based on his experiences with the Answer Garden projects, Mark mentioned the tremendous challenges (or pitfalls) involved in studying social effects within a software system. i.e.

1. Making sure the software developed is reasonable functional, useable and robust (at least the features we expect to evaluate). Users can be very unforgiving if we provide buggy software. They will simply not use it long enough to assess the social effects we are looking for.
2. The system must be somehow integrated within their existing social/daily pattern of tasks, such that they would naturally wish to utilize it in an ongoing basis.
3. One must have a clear hypothesis to start with (even if it must be revised later), to guide the sort of data one is looking for; else one would spend a great deal of time gathering an excessive amount of redundant data that would be hard to analyze/utilize.
4. Even during usage, users may continue to request feature changes or bug fixes, which may completely overwhelm the researcher (if he is the lead developer of the system as well). Monitoring usage, collecting data, interviewing users and maintaining a large software system is a tall order to accomplish together in a few short months.
5. You absolutely must know what data you will need and figure out how to make sure you get it. Even if one relies on interviews or usage data, in case either becomes hard to get, one should at least have some semi-structured questionnaires or other comments to fall back on. You may want a third-party to collect data that requires a random sampling or scheduling (e.g., critical incident interviews).

***Bish Sanyal, DUSP, MIT, Dec 19th, 2001***

Bish believes we ought to ask questions that counter the current assumptions.

The central issue is that of intellectual property rights. Who owns what and contributed what in the cooperative process? The question of understanding conflicts in the cooperative process and how they are resolved is most important.

Why would an online cooperation be any different? Consider what the literature says about how people have explained previous cooperation without such technology. How will that process be different now?

What does sharing mean? [Bish was likely referring to my proposed thesis title "Open Collaborative Design: *Sharing Evolving Design Knowledge in Distributed Social & Institutional Settings*"]. One must define it. It is not just "communication"!

Consider where you will get the data for your study. At the moment it seems like classroom projects will provide the bulk of the data. One must recognize that the learning and cooperation process is different in this case vs. typical group cooperation in the real world, where the constraints and conflicts will be more pronounced. The classroom projects may at best provide an idealized simulation of real-world cooperation. **In a classroom setting the stakes are not too high and real-life conflicts may not be observed.** For example, if monetary incentives and funding issues are not involved, the study/analysis should be framed to take that into account.

One must consider the nature of the educational system i.e. Technical school like MIT vs. a design school like Shristi. One needs to compare them and hold same aspects constant e.g. in India consider a design school like Shristi vs. technical school like IISC in Bangalore. Look at

differences in pedagogy and learning. It will affect the sorts of questions asked and patterns of learning/cooperation expected. The production of (cooperative) design in each institutional context will vary. Consider how the process of design is unfolding in each context; bottlenecks and sequencing of cooperation; codify the patterns of design or intellectual interaction. The relationship to NGOs and social sensitivity.

A PhD thesis can be broken up into three key chapters. Consider what those should be. The thesis defense or job talk must be able to outline these three aspects clearly. For example perhaps, 1) technological affordances, 2) patterns of learning/cooperation, and 3) institutional issues or incentives for participation.

The key issue for PhD research is the **methodology of research** conducted. The analysis should be motivated by goals, indicators and variables. Consider the counter indicators that would turn down the key hypothesis.

Be skeptical about all claims proposed. Consider what data would effectively disprove these claims. Test the underlying assumptions that lead to these claims. E.g. if claim is that open design leads to better learning, define "open design" – what is it? How can we test or refute this claim.

For any content analysis – if you conduct surveys you must make sure that the sample size is representative. For the few detailed interviews must select an appropriate sub-sample including people whose responses would vary with respect to the hypothesis. Must talk to all the different participants, for example particularly the NGOs who participated.

Some relevant readings include "The Reflective Practitioner" by Donald Schom and Chris Agary's work on collaboration and learning; he discusses "Conceptual Categories". Review Bish's paper on how sometimes conflicts in cooperative efforts have lead to better outcomes (slum housing).

**Prof. Anil Gupta, IIMA, Ahmedabad, India, Jan. 10-11, 2002**

Prof. Gupta suggested three aspects one could examine regarding "collaborative solving or solution augmentation".

1. *Behavioral* – cognitive issues and heuristics in collaboration.
2. *Computational* – tracking tools and expert systems.  
e.g. Intelligent search tools to find prior art using analogical or conceptual similarity (cited work of David Martin).
3. *Interface* – technologies for high-speed low-skill conditions.  
i.e. allowing everyday users to rapidly contribute/respond.  
He cited the INSTAR registration system for small-scale innovators that could provide low overhead means for registering and protecting their ideas.

Prof. Gupta asked me to carefully reconsider the core thesis assumptions:

- From an innovator's point of view, why would he join an open source collaborative? Without an appropriate protection or licensing mechanism (even GPL?), why would anyone give venture capital in the future?
- Collaboration has always occurred in many contexts. What is new about an open source approach to collaboration? What is the "malleability of technology" developed, implied with an open source approach?
- **Is "collaboration" even necessary or effective in every setting? Under what conditions?** What are the challenges in having folks want to collaborate? Perhaps the domain requires disparate expertise that cannot be harnessed without collaborative effort.

Prof. Gupta suggested that to verify the claims of collaborative design, one would have to construct an experiment with 2-3 different conditions:

E.g. Same Team (or different teams) collaborates in each of 3 conditions:

1. *Physical & Closed Loop* – i.e. face-to-face without external exposure of ongoing design.
2. *Virtual & Closed Loop* – i.e. face-to-face and distributed interaction within a small group of designers and peer-reviewers, with documentation online, but controlled access to the content and no public release.
3. *Virtual & Open Source* – i.e. interaction and archiving of ongoing design in the public domain.

*Is this a feasible setup? Would teams be willing to participate in such a setup? What would one want to measure or observe? What kind of outcomes would one expect? What is the role of computational tools?*

Prof. Gupta insisted that one needs to “kill open source” arguments first, before one can validate a suitable hypothesis for research. Some killing arguments he suggested:

1. Intellectual Property (IP) Protection gives inventors an incentive to disclose. However without IP, if they are secretive they don't generally collaborate.
2. What would prevent large companies from taking advantage of such an open system without contributing?
3. Smart people have an asymmetric incentive to participate if they feel they are not getting reciprocal benefits.

Generating Reciprocity is a real challenge. What are incentives for reciprocity?

Prof. Gupta mentioned 2 kinds of reciprocity: specific (immediate and equal return) vs. generalized (unconditional with no equivalence expected). He mentioned a profound quote “Unredeemed IOUs are the glue of relationships”.

One needs to understand all barriers involved in an open source culture. What about the issue of anonymity – how do people provide anonymous contributions or do they?

Prof. Gupta mentions a paper by Benson that outlines 4 strategies for resolving problems – Disruptive, Authoritative, Manipulative and Collaborative. What are the conditions where collaboration becomes ideal?

A reference was made to work by Charles Perot (MIT economist?) on his radical theories about small firm networks and the cycle of competition and collaboration observed over time (small textile or manufacturing firms in Northern Italy?). Another paper “Are solutions the problem?” also helps one change the frame of thinking. Finally a paper “Camping on Seesaws” by Starbuck and Nistron? (1978-79) talks about how resource constrained situations lead to better innovation.

Hence several modes of collaboration can exist. One needs to create tools for problem solving, and not necessarily just for collaboration as the prime goal. The platform should make it possible for individual as well as team problem solving & contributions.

**Key take-home messages from committee:**

- The need for “Collaboration” should not be prematurely presumed in every setting/project. Participants have different incentives and motivations to work independently or collaborate. There is probably a complex set of inter-related factors that lead to effective collaborative initiatives, particularly with distributed participants. So one cannot have blind faith in people's desire to collaborate as a pretext for research.
- One has to be careful what one can hope to learn and generalize from classroom-based cooperative projects; the economic stakes and conflicts in the real world are different.
- The methodology of research is more critical than necessarily the outcomes.
- Deploying a system and studying it is a tall order; one has to be confident that both the system and community of users are well in place if one wants to study relevant social aspects. One must be absolutely clear about what sort of data and social effects we are looking for and how to get it, even if things don't go as planned with potential users.