International Conference on Information Systems (ICIS)

ICIS 2009 Proceedings

Association for Information Systems *Year* 2009

Matching Mechanisms to Situations Through the Wisdom of the Crowd

Jeffrey V. Nickerson* Doris Zahner† James E. Corter‡
Barbara Tversky** Lixiu Yu†† Yun Jin Rho‡‡

*Stevens Institute of Technology, jnickerson@stevens.edu
†Stevens Institute of Technology, dzahner@stevens.edu
‡Columbia University, jec34@columbia.edu
**Columbia University, bt2158@columbia.edu
††Stevens Institute of Technology, lyu3@stevens.edu
‡‡Columbia University, yjr2101@columbia.edu

MATCHING MECHANISMS TO SITUATIONS THROUGH THE WISDOM OF THE CROWD

Jeffrey V. Nickerson
Stevens Institute of Technology
Hoboken, New Jersey
jnickerson@stevens.edu

Doris Zahner
Stevens Institute of Technology
Hoboken, New Jersey
dzahner@stevens.edu

James E. Corter
Teachers College, Columbia University
New York, New York
jec34@columbia.edu

Barbara Tversky
Teachers College, Columbia University
New York, New York
bt2158@columbia.edu

Lixiu Yu
Stevens Institute of Technology
Hoboken, New Jersey
lyu3@stevens.edu

Yun Jin Rho
Teachers College, Columbia University
New York, New York
yjr2101@columbia.edu

Abstract

Designing a system often begins with matching existing solutions to current problems. Specifically, integration mechanisms are mapped onto situations. Novices are not good at this task, and experts are rare. Could crowdsourcing, that is, aggregating the suggestions of individuals working independently, be effective? Two experiments, one with design students in a classroom, and another with participants on the web, demonstrated that the crowd possesses wisdom about how to match mechanisms to situations. Participants also categorized situations, and those who name their categories were better at matching than those who didn't. The results have pragmatic implications, suggesting it is possible to crowdsource design, and providing new ways of eliciting, testing, and training expertise. More generally, the paper suggests a new model for information system design based on analogical mapping.

Keywords: information systems design, visualization, categorization, analogical mapping, wisdom of the crowd, integration mechanisms

Introduction

The surgeon knows when to reach for the splint, and when to reach for the scalpel; the mechanic, the pliers and the wrench. Architects know where to put a porch and where to put a courtyard. It is clear that expertise involves matching solutions to problems, mechanisms to situations (Larkin et al. 1980; Chi et al. 1981).

For example, information systems integrators often design systems by matching previously proven solutions to new problems (Bergman et al. 2001; 2002). Software engineers also perform such pattern matching (Gamma et al. 1995); so too business process designers (Malone et al. 2003) and scientists of many stripes (Hamming 1986). But although it seems obvious that problem solvers would effectively transfer old solutions to new problems, they often fail to do so (e.g. Cohen et al. 1972).

One reason that designers fail to recall relevant old solutions to new problems is that they do not see the connections between the two; that is, they fail to represent problem and solution in the same terms, so that they would remind the designer of each other. This is often because problems are represented using surface structure, specific to the domain (Holyoak and Koh 1987; Schank 1982). Solutions are usually represented more abstractly, independent of domain. Becoming an expert is in large part a process of abstraction, of learning to represent problems free of domain content. It is also a process of learning solutions (Simon and Simon 1978; Sweller 1988). But it is hard to train people to properly apply ideas learned in one context to another context (Gick and Holyoak 1980; Novick 1990; Novick and Hmelo 1994; Ross and Kennedy 1990). Training is not a ready substitute for experience (Larkin et al. 1980); learning a profession is a slow process (Schön 1983). Indeed, expertise can be predicted by the time spent in deliberate practice: according to some, it takes 10,000 hours to master most skills (Ericsson et al. 1993; Simon and Chase 1973). This is discouraging for the field of information systems, because the chaotic nature of work in high tech often leads to high turnover (cf. Isler 2007). Thus, the prognosis is bleak: Design expertise takes a long time to develop in a rapidly changing technical context, and, once acquired, may be too rare or rarified to be influential. What can be done?

Because expertise is both rare and desirable, we propose that crowd-sourcing might serve as an alternative to the lengthy training needed to develop experts. While a particular individual may suffer from a variety of short-comings – myopia, lacunae, lapses, interference, biases – aggregated matches may work better than individual matches because the errors of human problem solving processes will cancel out. Thus, the wisdom of the crowd is often greater than the wisdom of any particular individual (cf. Dalkey and Helmer 1963; Galton 1907; Gurnee 1937; Kittur et al. 2007; Surowiecki 2004; Wagner and Back 2008).

We report two experiments that demonstrate that there is wisdom in the crowd. The results suggest that at least one component of information systems design, the matching of mechanisms to situations, might be accomplished by aggregating individual's responses. We thereby illustrate a way to design through crowdsourcing, based on the group's aggregated analogical mapping to information systems. These results suggest new ways of eliciting, testing, and training expertise in information systems design.

In the next section, we propose a theory of analogical mapping that we use to make specific predictions about the design expertise of individuals in relationship to crowds. These predictions are tested, first in a classroom situation, and then in an online web-based experiment with a diverse population of novices and experts.

A Theory of Analogical Mapping in Information Systems

As part of transitioning from novice to expert, a designer masters the ability to find the most appropriate known integration mechanism to fit a particular design problem. We call this process *matching*.

Let us imagine two designers, novice and expert, given the task of constructing a reservation system for a theatre. The novice is confronted with the problem of representing the problem correctly. The novice must distinguish between superficial aspects of the problem (how expensive are the tickets?) and important ones (are specific seats reserved?). The novice might flounder in the large solution space suggested by the many aspects of the problem, asking: should this be a workflow implementation? Part of a publish/subscribe system? Or can it be done simply through email? Indeed, novices suggested all of these unsuitable ideas in our first experiment.

The expert designer, in contrast, knows how to represent the problem. Hence, the designer can narrow the solution space. For example, the expert might reason in the following way. This particular theatre has assigned seats, so before buying tickets, customers will want to know their seat assignments. Moreover, it is quite possible that multiple people are seeking tickets at about the same time. Therefore, a technology is needed that will inform the reservation agent which seats are available, at the same time preventing these seats from being taken by other customers until the present customer's decision is completed. For this, a locking mechanism is needed which temporarily blocks particular seat assignment once these seats have been offered. Locking mechanisms are built into transactional databases, and so viable solutions include buying a system that includes a transactional database, or building a customized application on top of a transactional database.

How do we produce this expertise? We know that, in general, experts make decisions better and faster than novices. They abstract problems and solutions, and they chunk both into manageable groups, drawing on a larger set of experiences (Larkin et al. 1980; Simon and Chase 1973). From this it is easy to conclude that focused, deliberate practice is the prerequisite to expertise (Ericsson et al. 1993). That is, the way to train an information systems designer is to apprentice that designer to a senior designer, who will function as a tutor, providing a set of practice exercises over many years. In many disciplines, this is effectively how design is taught (Schön 1983).

But perhaps there is another way of thinking about the matching problem. Along with others, Holyoak and Thagard (2002) show that many problems are solved through analogy. They borrow the tumor problem from Duncker (1945): a doctor needs to destroy a tumor; there is a special ray that can do it, but the amount of energy needed to kill the tumor will also kill any intervening tissue. The solution is weaker rays that converge on the tumor from different angles. This convergence problem has many analogs. For example, soldiers may converge on a fortress from different directions, as any one street will not hold enough of them to afford critical mass in the battle. While the process of making such an analogy seems mysterious, Holyoak and others (e.g. Gentner and Markman 2006; Gick and Holyoak 1980; 1983; Goldstone and Sakamoto 2003) have built models of these processes that can replicate human performance in certain domains. These models work on the concept of *structural similarity* (Gentner and Markman, 2006). That is, even though the surface similarity between a medical procedure and an attack on a city is low, at a deeper, structural, level, the situations share certain characteristics. Furthermore, analogy can work at different levels of abstraction: it is possible to recognize the medical problem as just an example of a more abstract convergence problem. To return to our example, the expert recognizes the theater reservation problem as one that involves temporary exclusive protection of a resource (locking); that in turn suggests a technological mechanism, a transactional database. Generally, the problem and the solution will be semantically dissimilar, but structurally similar, and the expert learns to recognize this structural similarity.

Thus, the design matching problem in information systems seems to be well modeled as analogical reasoning, based on the structural similarity between a problem (a situation or set of requirements) and a solution (an integration mechanism or technology). Analogical reasoning is difficult because so much must be mastered: the problem and solution domains, the elicitation of relevant features, the abstraction of unimportant surface detail, the simultaneous matching of related features in both domains (cf. Gick and Holyoak 1980, 1983). What's more, in information systems design, the analogies are rarely perfect, so the designer must decide which aspects are critical, and which are not, and to improvise. It is understandable that novices are swamped, and why experts make mistakes. But because so many factors are involved, errors on such tasks are likely to be uncorrelated in a crowd that is diverse with respect to experience. Thus, crowd-based aggregation techniques may be particularly suited to such matching problems.

Crowd-sourcing techniques long predate the name: Galton (1907) found that the median of guesses about the weight of an ox at a fair was within one percent of the actual weight. Galton observed that those guessing about the weight varied in expertise: the population included passersby as well as farmers. Moreover, the median guess was better than anyone's guess. Similar results involving weight estimates have been obtained in carefully controlled experiments (Gordon 1924; Bruce 1935), and in a variety of other tasks (Lorge et al. 1958). Notably, Gurnee (1937) showed that a group could learn their way through a maze faster than individuals: they moved by voting on each turn, and following the mode.

While many of these experiments were bound by the restrictions of geography, the Internet provides an infrastructure that makes it possible to easily solicit and aggregate ideas across space and time (Wagner and Back 2008): The term *crowdsourcing* (Howe 2006) has been coined to describe this activity in both academia (e.g. von Ahn 2006) and industry (e.g. Wessel 2007). The work of Gurnee and others, combined with presence of Internet infrastructure, suggest that certain problems can be readily solved through the aggregation of independent individual decisions. Here we propose that the matching problem can be accomplished effectively through the wisdom of the crowd.

We know that experts chunk things differently from novices, at least in domains such as chess playing (Simon and Chase 1973). Is this also true in information systems design? In addition to asking designers to solve the design matching problem, we also ask participants to categorize mechanisms and situations, so that we might better understand the aggregated view of the crowd; that is, the view of the *statisticized group* (cf. Knight 1921; Lorge et al. 1958). In our previous work, we have described how to build an aggregated sketch — a consensus graph — of an information system (Nickerson et al. 2008). We showed that many highly varying solutions could be synthesized into one strong solution. Our focus, however, was on just one specific problem. Can this idea be used more generally, to choose mechanisms for a wide variety of problems? We answer this question by applying our technique not to participants' solutions to a single design problem, but instead to participants' matchings between many problems and many solutions. We call the result a *consensus matching* of situations and mechanisms. Our previous work showed that the consensus graph appeared to be the product of expertise: we will look for the same phenomenon here.

While we think crowdsourcing will be an important contributor to future design activity, as educators we also want to increase individual performance, moving students from novice to expert as quickly as possible. Therefore, we will also test to see if the way an individual categorizes situations or mechanisms can help us predict matching performance. With respect to such categorization behavior, studies have shown that *naming* improves the ability to attend to the right properties of objects (Smith et al. 2002). The names themselves may not matter: category learning can be improved just by providing nonsense labels for the categories (Lupyan et al. 2007). Labels are themselves an abstraction, and using them has been found to increase overall abstraction (Tversky et al. in press). The ability to abstract is a prerequisite for successful analogical reasoning (Holyoak and Koh 1987). Thus we will test whether labeling of categories is correlated to success in the matching task.

In sum, we see the matching problem as a form of analogical reasoning. We predict that the crowd, constituted by taking the mode of the designers' choices, will be better than the average individual designer at the matching problem. We also predict that the consensus matchings, the graphs representing the group view of matches, mechanisms, and situations, will exhibit signs of expertise. Finally, we predict that individual categorization skills will relate to the individual's ability to match situations and mechanisms – specifically, that the labeling of categories will correlate with matching scores.

An Experiment in Matching and Categorizing

Thirty-seven students from a Master's level information systems design course (cf. Nickerson 2006) participated in this study during the fall semester of 2008. Participants had varying levels of expertise: some students were novices, just beginning their technical training, while other students were working professionals with five years of experience in systems design activities.

Participants were presented with a list of thirty-eight integration mechanisms – including *using a publish/subscribe system, sending an email, text messaging, teaching in a classroom, teaching online, creating a television broadcast, leaving a voicemail* – and forty-five situations – including *finding a job, backing up information in anticipation of a hurricane, distributing stock quotes to traders, announcing a lecture, informing students of an outbreak of measles*. These mechanisms and situations were chosen to span a wide range of technologies and situations that occur in practice.

The participants were asked to "pick the integration mechanism or mechanisms" they thought were appropriate for each of the 45 situations. If they thought more than one mechanism was appropriate, they were asked to order them with the most appropriate mechanism listed first. Following the matching task, the participants were asked to categorize the different mechanisms by grouping them together "in whatever way you think makes sense". Then they were asked to categorize the situations the situations in the same way.

From the matching task, a matrix of the matched situations and integration mechanisms was created for each participant. Each situation was then analyzed to compile the matched integration mechanisms, which were ranked in order of frequency in the group listing. In total, 37 matrices were created. Additionally, an expert with twenty years of design experience was asked to complete the same task, resulting in a 38th matrix. Abbreviated descriptions of the situations and modal mechanisms are given in Table 1.

We calculated the wisdom of the crowd by choosing the modal answer for each question. For analysis purposes, we considered the expert's answers to be correct, and measured the proportion of matching to the expert answer for both the individuals' solutions and the crowd's modal solution. The results are shown in Figure 1. As predicted, the crowd matching score is higher than the mean score of the individuals.

We noticed that sometimes there was strong agreement matching a mechanism to a situation, and other times multiple mechanisms were picked. Was this the result of a small sample size, or was it the nature of the particular situation? In other words, are these situations ones for which designers will tend to find many appropriate mechanisms? We tested this by taking one of these low-agreement situations, the presentation of a job offer, and presenting it as a multiple choice question to a different sample of fifty participants, this group from an online population. The mechanisms chosen were split much as they had been during the in-class study, suggesting that some situations have a dominant solution, and others have several viable alternatives.

Table 1. The most commonly chosen mechanism (the mode) for each situation with the proportion of participants choosing the mode.

Situation	Mechanism	Proportion
1: job	D: social network	0.526
2: branch offices	F: VPN	0.368
3: online customer feedback	FF: online discussion board	0.263
4: db merger	B: consolidated db	0.553
5: backup info	W: db replication	0.658
6: stock quotes	C: pub/sub	0.711
7: lecture time	I: email	0.526
8: measles	I: email	0.368
9: imminent threat	M: megaphone	0.421
10: expense account approval	J: workflow	0.263
11: physical records offsite	T: courier service	0.421
12: seat reservations	A: transactional db	0.395
13: mentoring	S: face-to-face	0.500
14: sending large file	BB: thumbdrive	0.500
15: restaurant	Y: physical exchange	0.316
16: network PCs	E: Ethernet	0.842
17: side effects of drug	H: search engine	0.763
18: discuss urgent matter	Q: telephone	0.763
19: contact college friend	GG: facebook	0.605
20: build open source app	FF: online discussion board	0.237
21: update friends w/out interruption	I: email	0.395
22: least loaded db server	O: data network broadcast	0.395
23: send daily large report electronically	P: FTP	0.500
24: remote laptop info reaches HQ	F: VPN	0.289
25: minutes from board meeting	V: physical conference	0.763
26: presentation	AA: pub website	0.289
27: consumer products	L: TV broadcast	0.316
28: trend	AA: pub website	0.316
29: new student orientation	EE: WIKI	0.289
30: teach driving	HH: tutoring	0.711
31: teach accounting	II: teaching classroom	0.553
32: teach design	II: teaching classroom	0.711
33: current trend discussion	FF: online discussion board	0.605
34: political news	KK: pub in newspaper	0.316
35: Call in sick	I: email	0.447
36: work harder	S: face-to-face	0.737
37: job offer	LL: writing letter	0.211
38: lay off	S: face-to-face	0.711
39: convert files	N: web service	0.342
40: coordinate meeting	I: email	0.526
41: favorite author's next book	C: pub/sub	0.711
42: deposit slip	A: transactional db	0.816
43: 10% discount offer	I: email	0.289
44: interoffice mail	Y: physical exchange	0.421
45: coordinate emergency response	Q: telephone	0.211

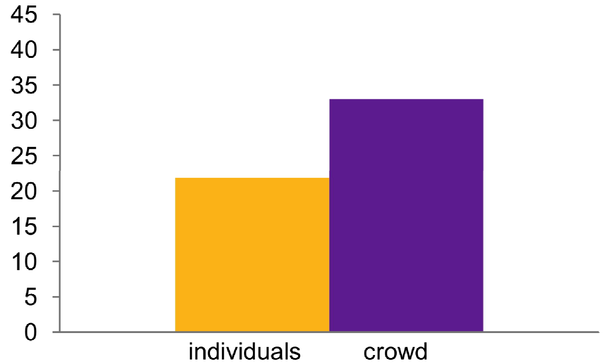

Figure 1. Mean number of integration mechanism matches to expert's solution for individuals versus for the crowd's modal mechanism.

An expert looked at the modal answers for the matching exercise, and found that all the answers seemed reasonable. How does the crowd compare to the expert? We compared the expert's choices with the modal integration mechanism for each situation. As shown in Figure 1, the crowd and the expert agreed on 33 out of the 45 modal integration mechanisms. Since our results suggest that for some problems, several alternatives are appropriate, we looked at individual solutions for each problem and counted the number of participants who had at least one of the same answers as the expert on that problem, even if it was not their first choice. Using this analysis, we found that the expert agreed with the crowd 37 out of 45 times.

There were eight situations in which the expert's choice was distinct from the crowd's choice. Perhaps the expert was just idiosyncratic? We asked two other experts to choose and rank the best integration mechanisms for those eight situations in question. One of the new experts agreed with the original expert in seven out of the eight situations and agreed with the crowd on one of the situations, while another expert agreed with the original expert on six out of the eight situations, agreed with the crowd on one of the situations, and agreed with neither the original expert or the crowd on the last situation. Thus, even though experts often disagree (Shanteau 2000), here it seems the experts share similar opinions. However, it is not clear the experts are really wiser than the crowd: one expert noted that crowd's answers were appropriate, and might represent generational differences in attitudes toward technology.

Now we turn to look at the categorization task. Students were asked to cluster both the mechanisms and the situations. No category labels were supplied. The diagrams in Figures 2 and 3 show the variety of categories used for mechanisms. In general, students chose to split the set into a small number of categories, usually two, and used Venn diagrams to classify the items within the categories.

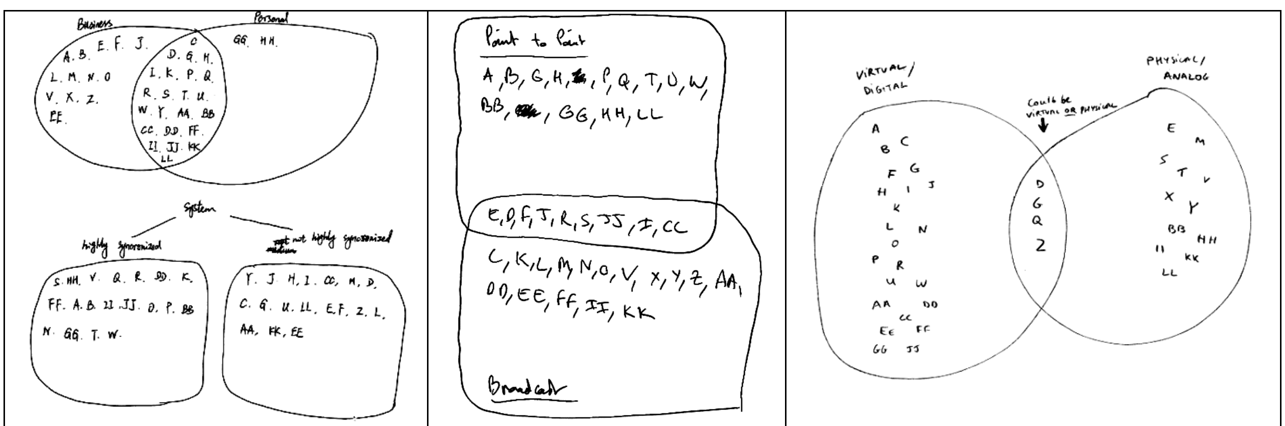

Figure 2. Three examples of the use of Venn diagrams to categorize mechanisms.

For example, in Figure 2, one participant, in the graph on the left, first broke the mechanisms into those appropriate for business, for personal use, or both. The same participant then also classified the mechanism according to how synchronized they were, differentiating between highly synchronized and not highly synchronized. A different participant, in the middle diagram, split the mechanisms into point-to-point and broadcast technologies. A third participant focused on the mechanisms as being virtual/digital versus physical/analog.

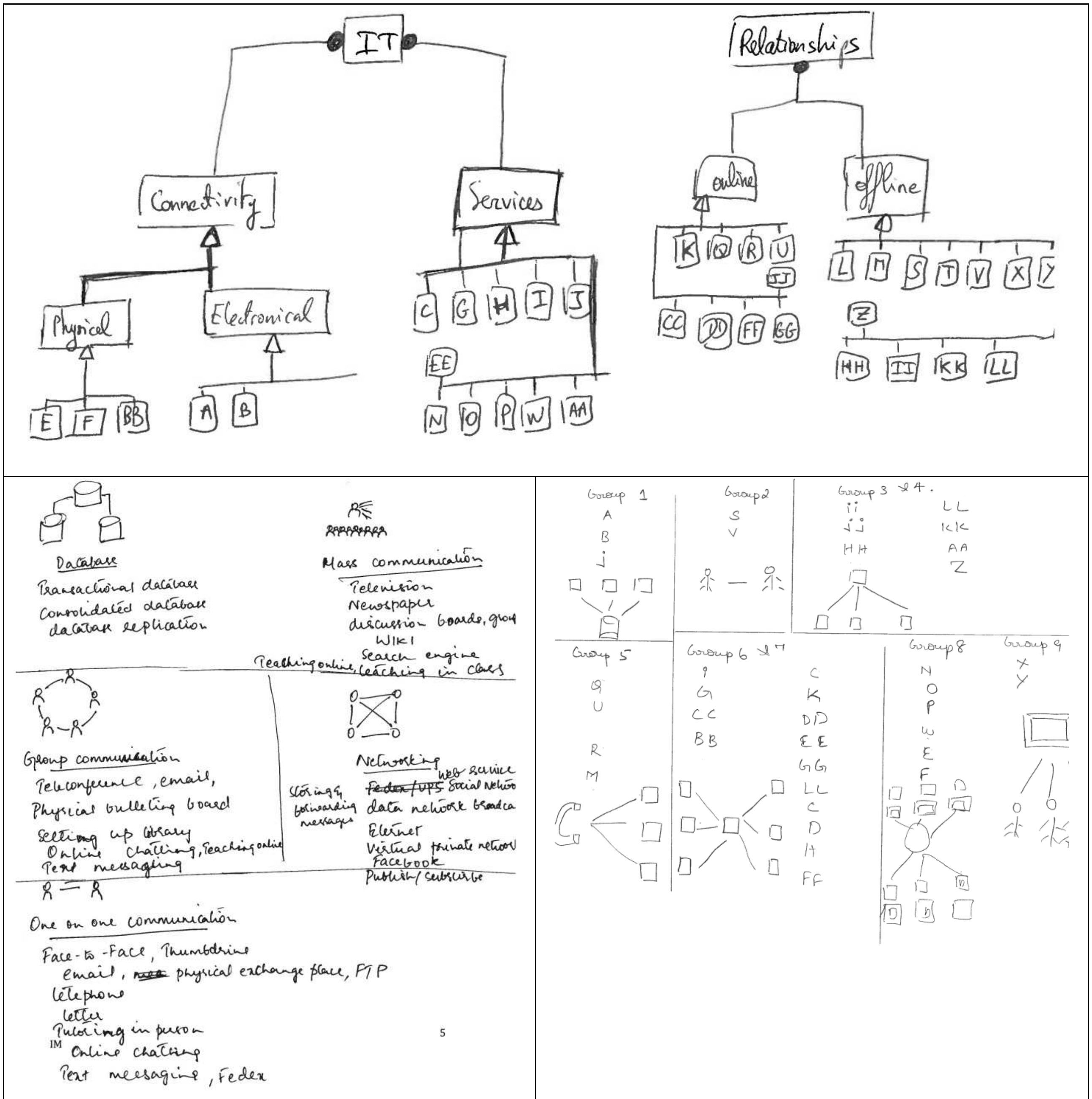

Figure 3. Other integration mechanisms examples. The top diagram is a set of tree structures designed by one participant. The bottom two diagrams are categories created by two separate participants.

More complex schemes are shown in Figure 3. At the top, and designer first split the mechanisms into the categories *IT* and *Relationships*. Then this designer further subdivided the IT portion into *Connectivity* and *Services,* with *Connectivity* being further subdivided into *Physical* and *Electronical*. Under the box labeled *Relationships*, the participant differentiated between *online* and *offline* mechanisms. A second participant split the mechanisms up into five categories: *database*, *mass communication*, *group communication*, *networking* and *one-to-one communication*. A third participant provided a similar breakdown to that of the second: the categories are not labeled with words, but instead with diagrams suggesting different forms of communication.

From this extremely varied set of categories, can we form a group consensus? We performed the analysis in the following way. Mechanisms or situations were coded as belonging to the same category if the participant grouped them together by drawing lines around them or spatially separating them as in Figures 2 and 3. Each item in the intersection of the Venn diagrams in Figure 2 was coded as belonging to multiple categories. More generally, items could occur in multiple categories, through Venn diagrams or through multiple listing, as in the left diagram of Figure 2: Mechanism A (a transactional database) was coded in both the category *business* and the category *highly synchronized*.

For every student, a graph was formed. If any two mechanisms occurred in the same category, then an edge joined them. Thus, a participant-defined category formed a completely connected subgraph (cf. Harary 1994) of mechanisms or situations. We took all of the student graphs and combined the edges, weighting the edges according to how many times these edges occurred in the graphs. Complete details of the graph transformations are explained in Nickerson et al. (2008). Figure 4 displays the resulting graph using a spring-electrical embedding (Hu 2006), adjusted to enhance the visibility of the labels.

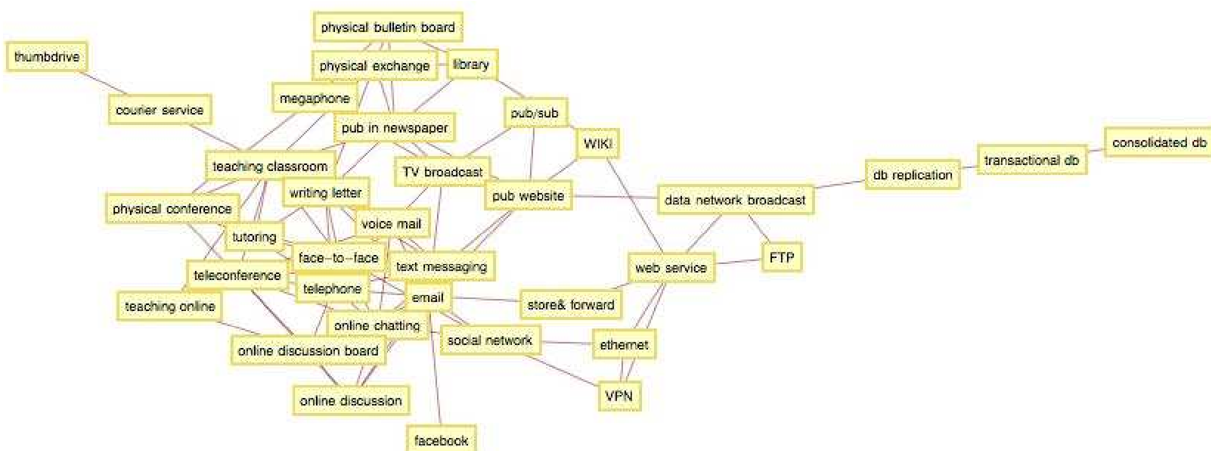

Figure 4. Consensus graph for the integration mechanisms

There are several interesting aspects of this graph. On the left top corner, a thumb drive and courier service are related; these are two techniques that involve moving physical objects. On the far right corner, all the database techniques are clustered. In the center of the graph is a dense cluster of commonly used techniques: email, face-to-face meetings, and online meetings. Above these are broadcast mechanisms: newspaper and television broadcast, as well as a megaphone. Some links are surprising and provocative: is teaching in a classroom really like a courier service? Or publishing in a newspaper? Students may find classroom settings less interactive than professors imagine them.

Overall, it appears that the clustering has wisdom in it; that the consensus graph represents a coherent view of integration mechanisms. What about the situations? Figure 5 also shows an intelligent structure. Up on the left, we see a number of situations related to the workplace. On the bottom right are situations in which data needs to be moved. It is possible that the situations are sometimes clustered on the basis of the mechanism. To investigate if this is true, we formed a bipartite graph in which the consensus relationship between integration mechanisms and situations is illustrated (Figure 6).

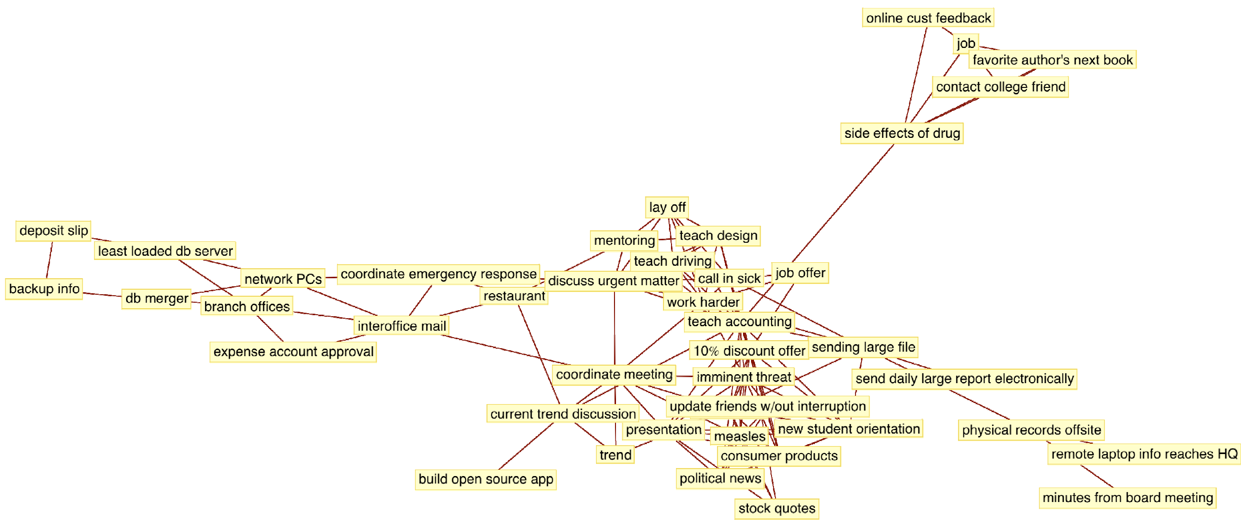

Figure 5. Consensus graph for the situations

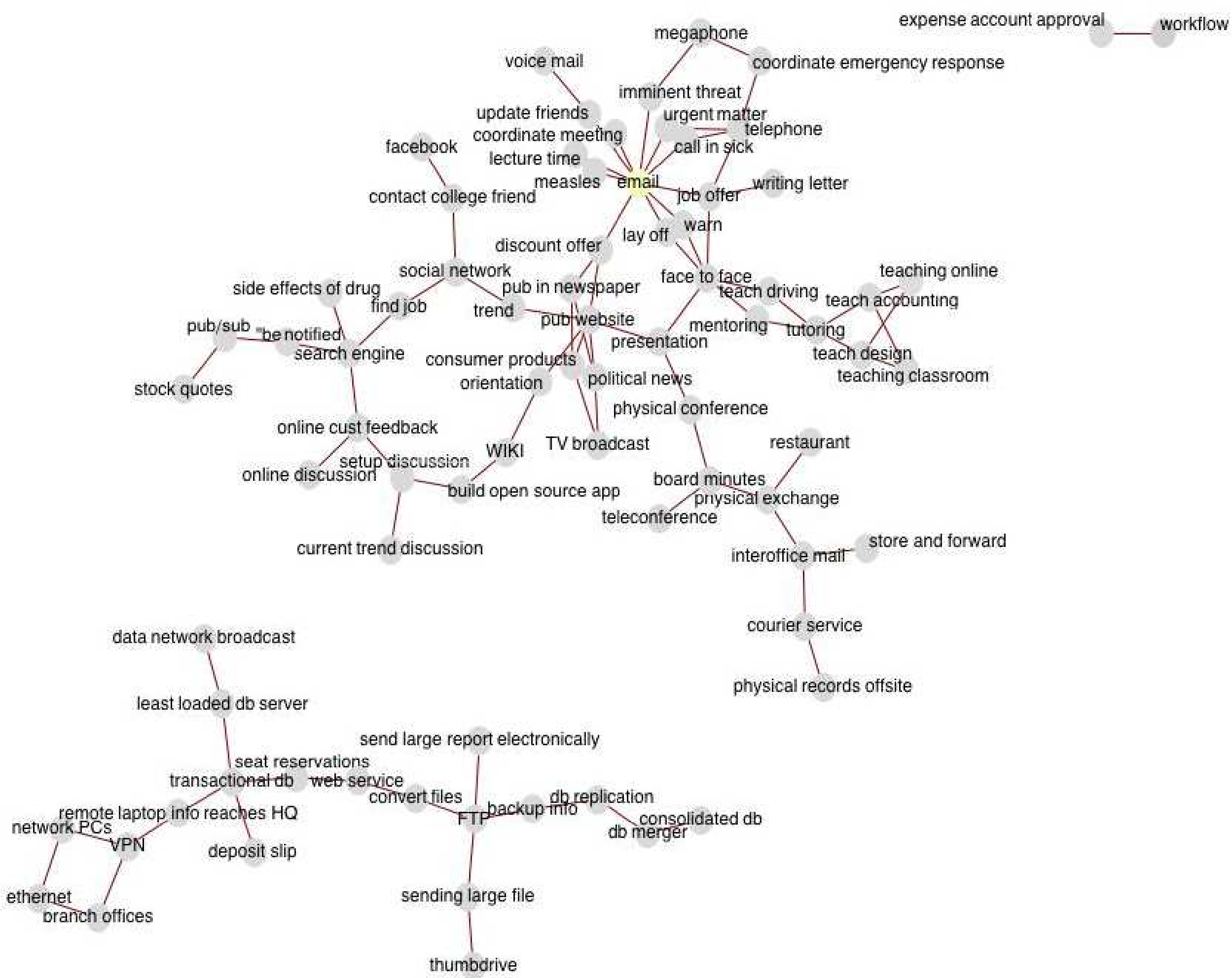

Figure 6. A bipartite graph of mechanisms and situations

An edge is shown between mechanisms and situations if at least five participants made the match. Notably, the graph breaks into three connected components. At the top center is the largest component, with *email* as a central hub, shown in yellow. At the bottom is a component made up of data-related situations and mechanisms. At the top right is a two-element component: Workflow is seen as being especially appropriate in the case of an expense-account approval situation. Within the top middle cluster, we see that the *side effect of a drug* is a piece of information that can be found through an Internet search, explaining its unusual placement in Figure 5. Also, we note that a megaphone, on the top of the diagram, can be used to coordinate an emergency, or to announce an imminent threat on campus, and there are alternatives: the telephone for coordination, email for announcements. There are three main hubs in the large cluster*: meeting face-to-face, sending an email,* and *publishing on a website.* These might be interpreted as representing different degrees of synchronicity. A situation such as *laying someone off* lies between two of them: *meeting face-to-face* was chosen the most appropriate for this task by about three quarters of the participants, but some chose email, an impersonal but increasingly used mechanisms for job-related announcements. Thus, the map of Figure 6 seems sensible, representing a reasonable matching of situations to mechanisms.

We looked to see how those who did well on the matching test classified the problems and the mechanisms. We found that strongly performing students did not share the same classification schemes. That is, some strong students picked two categories, others more; some concentrated on synchronicity, others on the cardinality of connections. There was a big difference, though, between those who matched well and those who didn't. The worst performers, categorized *without using labels for the categories*, as shown Figures 7a and 7b below, while participants with good matching scores labeled categories, as shown in Figures 8a and 8b.

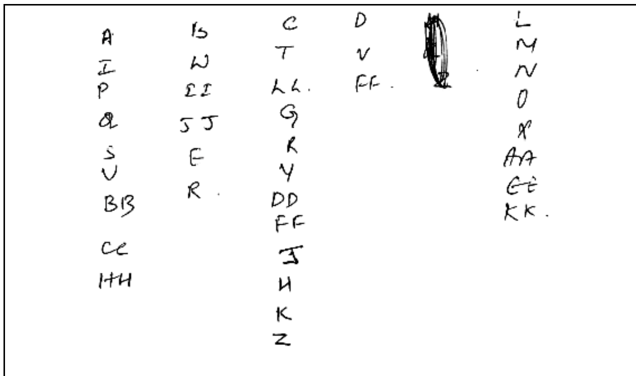

Figure 7a. An unlabeled clustering of mechanisms

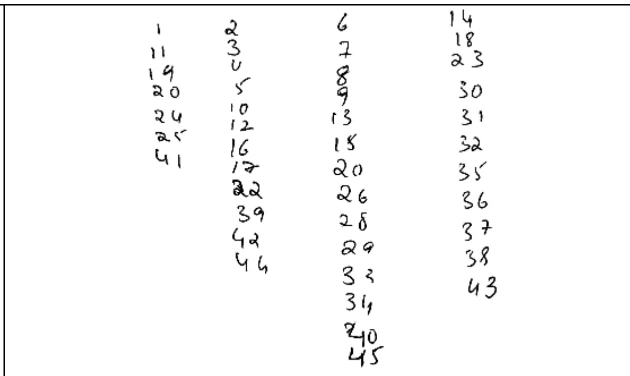

Figure 7b. The same person's clustering of situations

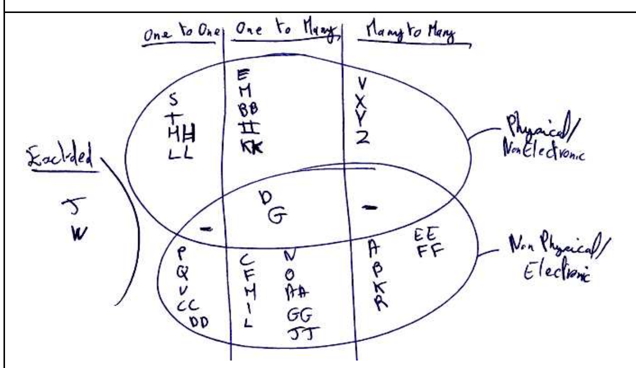

Figure 8a. Integration mechanisms groupings for someone with high matching scores.

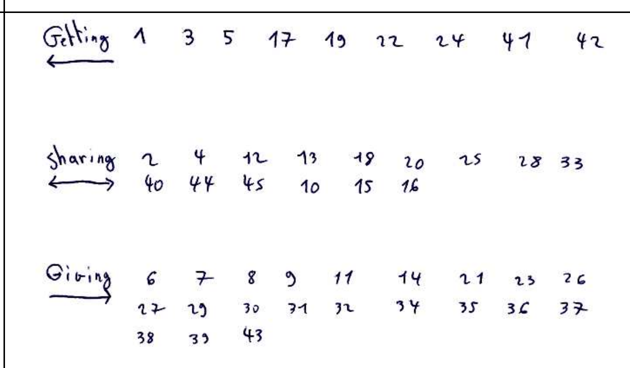

Figure 8b. The same person's clustering of situations

We tested this observation statistically: there was a significant difference between the matching scores of participants who labeled their clusters for the situations (M =23.32, s = 4.51) and those who did not (M= 16.57, s =

4.64), t(df = 35) = 3.51, p <.01. In other words, participants who used labels did much better: We show this in Figure 9.

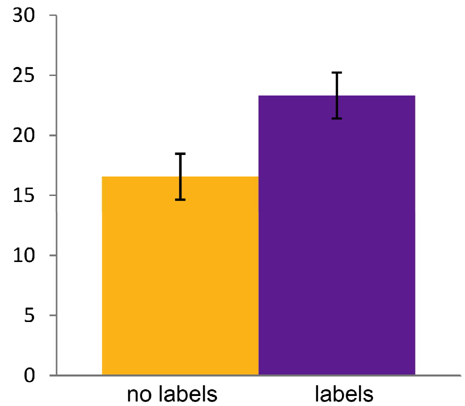

Figure 9. Mean number of matches; error bars represent standard error.

What might explain this? It is likely that those who name a cluster have abstracted the more pertinent features and are thus able to name them. These same students are likely the ones who find it easier to engage in the highly abstract reasoning involved in matching mechanisms to situations.

An Web-based Experiment

This experiment was designed to test if a different population of participants would also exhibit wisdom of the crowd with respect to matching. Specifically, we wanted to see if a different set of participants, outside of a classroom environment, would also do better as a group than the average individual, so we recruited participants from an online forum.

We asked participants to answer a set of six multiple-choice questions matching mechanisms to situations. We predicted that, as before, the modal answer would be correct more often than the average individual's answers. Six of the original 45 situations were used in this experiment. They included questions on: *the combination of the databases of two companies after a merger, distributing stock quotes to traders, creating a system for the approval of expense accounts, handling seat reservations for a theatre, finding out when your favorite author is publishing a new book*, and *discovering which database server is the most lightly loaded on the network.* For each situation, among the listed choices there was only one correct integration mechanism, as agreed by experts.

Participants were solicited through a posting on a public website asking them to *"Answer a set of 6 multiple choice design questions (knowledge of information systems is important)"*. Once they agreed to participate in the study, they were presented with instructions that told them to *"Imagine you are an information systems consultant presented with the following problems. For each problem, please choose the best technology to solve the problem from the set of choices provided"* followed by six four-choice multiple-choice questions. Ninety-three subjects participated in this study, 40 females and 53 males, and all were compensated with a nominal stipend. Their ages ranged from 17 to 69 with an average age of 32. They spent an average of 2 minutes 33 seconds completing the task and a related demographic survey. Seventy-seven percent were primarily English speakers and 67 percent had college degrees. Twenty-four percent had more than nominal programming experience (more than 10,000 lines of code) and 10 percent had more than five years of work experience.

For each question, we found the most frequent answer choice and compared that to the experts' correct answer. We then calculated the average number of correct solutions for the participants. Once again, we found that that the crowd as a whole did better than the average participant. The modal score for the task was five, meaning that for five out of the six questions, the modal answer choice was the correct answer. In contrast, the mean individual score was 3.22, meaning that participants on average answered between three and four questions correctly, as shown in Figure 10. Thus, we were able to replicate our results in an online setting conducive to crowdsourcing.

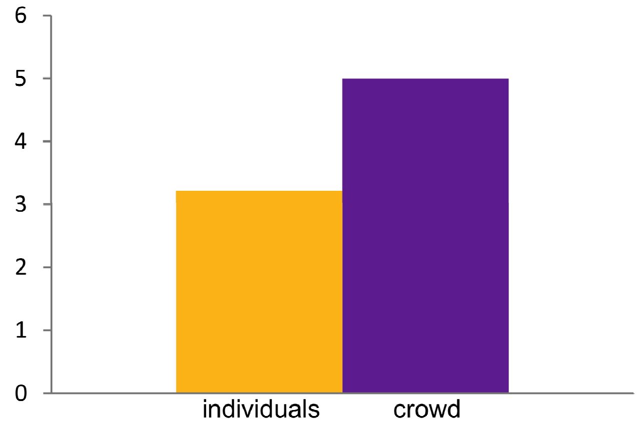

Figure 10. Mean number of correct solutions for individuals versus the crowd's modal number of correct solutions

Discussion

Why does the crowd exhibit such clear expertise? Perhaps the process of matching mechanisms to situations is one that requires broad experience: a particular individual may not have had the relevant experience for one specific question, but collectively the individual's experiences reinforce each other. Also, it may be that this matching task requires a high level of attention: individual's attention may lapse, but these errors presumably occur independently, so that the crowd is less likely to make common mistakes. Future work might disentangle these possible explanations by first varying the experiential diversity of the designers. For example, we expect that if the crowd were both homogeneous and inexperienced, then the crowd would not do much better than do individuals. Finding the limits to the effectiveness of crowdsourcing is an important future project, and might be accomplished by, on the one hand, testing a variety of different online populations, and on the other hand, attempting to figure out how matching choices are made, and how errors occur. Errors might be studied by increasing attention load in experimental conditions, to differentiate errors caused by lack of focus from those caused by lack of experience.

Why did students who named their clusters do better at a matching task than students who didn't? There are several possible explanations for this. First, it is possible that a lack of labeling is a symptom of lack of motivation. Alternatively, perhaps because students don't understand the mechanisms and situations, they can't articulate a category name. Lastly, the causality may run the other way: because participants haven't named things, they are less able or less prone to abstract, which hinders the analogical mapping process. In order to test these alternate explanations, an experiment might be run in which one group of students is asked to name their clusters, and another group is asked to cluster without providing names. Alternatively, students might be asked to explain their clustering choices to another student or to an experimenter, and their performance compared to students who are not asked to do this.

What is it that makes this matching task so difficult? While most analogical reasoning compares similar objects or similar situations, the matching of mechanisms to situations is more complex. A database is not at all similar on the surface to a theatre, but a database has features, such as the atomicity of transactions, that fulfills a requirement of theatres, the need to assign each seat to just one customer. Holyoak and Thagard (2002) model analogical reasoning as constraint solving: applying this idea, we can reason that an important constraint of a theatre is that only one person can sit in one seat during a performance, and, by extending this constraint to the reservation process, we infer the need to lock database records. It seems like a tall order to expect all novice designers to figure this out, because it took the software industry many years to identify the abstract situation called a *race condition* and construct its solution through the abstract mechanism called an *atomic transaction* (cf. Ullman 1988). Certainly, then, designers should be taught such abstractions. For the designer, though, there is a great deal of such material that needs to be mastered, because information systems typically operate on different modular layers, and appropriate mechanisms for solving a problem may exist in the network, operating system, middleware, database, application, user interface, or organizational layers. Moreover, new technologies are constantly appearing: at this moment, virtual machine technologies are changing the principles of enterprise architecture (Smith and Nair 2005), and electronically mediated social networks are changing notions of coordinated work (Benkler 2006). Thus, not only should abstractions be taught across many different technological layers, but also the ability to make new analogical mappings should be trained, because the universe of mechanisms and situations is constantly changing as new technologies supplant old, and economies shift the range of likely problems.

In order to teach such matching to students, we need to understand it better. The possibility of normative analysis might be explored first. Clark and Brennan (1991) suggest that we can assess when a technology fits a situation by looking at the costs and constraints of the technology. Sometimes, of course, these calculations will be hard to agree on, and participants in formal design evaluation processes will attempt to game the assessment process (cf. Kazman et al. 2001). Crowdsourcing might provide a way to decrease the effect of such gaming behavior. Besides a normative analysis, a descriptive approach might provide insights into actual, rather than ideal, design processes. This research might begin by collecting examples of expert behavior, as well as think-aloud transcripts of experts in action (cf. Ericsson 2005; Larkin et al. 1980).

What is the relationship of our findings to other work in Information Systems? While other theories have tried to measure task-technology fit (e.g. Goodhue and Thompson 1995; Zigurs and Buckland 1998), these theories focused on explaining user satisfaction. Instead, we are modeling the design process, with the goal of understanding and emulating the decision-making that leads expert designers to match mechanisms to situations better than novices. While users can and should be consulted in the early stages of design, it falls to designers to anticipate future situations and the appropriateness of integration mechanisms to those situations.

Where else might this research apply? In many domains problems and solutions need to be matched; this kind of matching is thematic, based on complementarity, rather than on surface similarity. Thematic association occurs in many professions: Doctors match procedures to illness, lawyers match precedents to cases, mathematicians match methods to problems. Thus work that uncovers the way expertise emerges, and how crowds exhibit expertise, may have broad implications in a number of fields.

Concluding Thoughts

The crowd does well at a problem of matching mechanisms to situations. This is likely because statistical measures of central tendency can be remarkably effective at extracting signal in a background of uncorrelated noise, as Galton (1907) recognized. This finding appears robust: Crowdsourcing worked both with a student population and with a general population. This result is encouraging, as it suggests new ways of generating and evaluating design.

There are many possible ways the work can be extended. Design has many intermingled stages: where might crowdsourcing be effectively used? Our previous work showed that it might be used to form solutions to specific design scenarios (Nickerson et al. 2008). Here we have shown that it might be used to find and match general mechanisms to classes of problem. Will crowdsourcing work with the refinement stage of design, or is it most effective in the early stages of design? This question can be investigated through a series of experiments in which crowds compete with experts in different phases of the design process.

The expertise of designers in a crowd doubtless plays a role in the success of this and related experiments. It seems unlikely that a pool of total novices will do as well as a pool with mixed expertise. Our student population, made up of both novices and part-time students with industry experience, had mixed expertise. Our online community was likewise mixed, but with a greater range of ages and experiences. We would like to better understand how sensitive the crowd is to the diversity of expertise: Does the crowd itself have a level of expertise that is a function of both the expertise of its members and its size? This might be studied through aggregating the responses of various subsets of a large crowd, and analyzing the subsets' consensus design results.

Pragmatically, our results suggest that design itself does not need to be the prerogative of lone designers or close-knit teams. Instead, design might profitably involve independent actors whose different viewpoints are aggregated. Online communities make this feasible. Thus, in the future, we might see information systems technologies emerge not just from the painful trial and error of the marketplace, but also from ideas synthesized from peer production networks, the wisdom of the crowd.

Acknowledgements

The authors gratefully acknowledge the support of the National Science Foundation (awards: IIS-0725223, IIS-0855995, REC-0440103) and the Stanford Regional Visualization and Analysis Center.

References

Benkler, Y. *Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press, New Haven, CT, 2006.

Bergman, M., King, J., and Lyytinen, K. "Large Scale Requirements Analysis as Heterogeneous Engineering," *Scandinavian Journal of Information Systems*, (12:1), 2001, pp. 37-55.

Bergman, M., King, J., and Lyytinen, K. "Large Scale Requirements Analysis Revisited: The Need for Understanding the Political Ecology of Requirements Engineering," *Requirements Engineering Journal*, (7:3),2002, pp. 152-171.

Bruce, R. "Group judgments in the fields of lifted weights and visual discrimination," *Journal of Psychology*, (1), 1935, pp. 117-121.

Chi, M.T.H., Feltovich, P.J. and Glaser, R. "Categorization and Representation of Physics Problems by Experts and Novices," *Cognitive Science,* (5:2), 1981, pp. 121-152.

Clark, H. and Brennan, S. "Grounding in Communication," in *Perspectives in Socially Shared Cognition* , Resnick, L.B., Levine, J.M., and Teasley, S.D. (Eds.), American Psychological Association, Washington, DC, 1991, pp.127-149.

Cohen, M.D., March, J.G., and Olsen, J.P. "A Garbage Can Model of Organizational Choice," *Administrative Science Quarterly,* (17:1), 1972, pp.1-25.

Dalkey, N. and Helmer, O. "An Experimental Application of the Delphi Method to the Use of Experts," *Management Science*, (9:3), 1963, pp. 458-467.

Duncker, K. "On Problem Solving," *Psychological Monographs*, (58:270), 1945.

Ericsson, K.A. "Recent Advances in Expertise Research: A Commentary on the Contributions to the Special Issue," *Applied Cognitive Psychology*, (19:2), 2005, pp. 233-241.

Ericsson, K.A. , Krampe, R. T., Tesch-Römer, C. "The Role of Deliberate Practice in the Acquisition of Expert Performance," *Psychological Review*, (100:3), 1993, pp. 363-406.

Galton, F. "Vox Populi," *Nature*, (75:1949), 1907, pp. 450-451.

Gamma, E., Helm, R., Johnson R., & Vlissides, J. *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley, Upper Saddle River, NJ, 1995.

Gentner, D. and Markman, A.B. "Defining Structural Similarity," *The Journal of Cognitive Science*, (6:1), 2006, 1-20.

Gick, M.L. and Holyoak, K.J. "Analogical problem solving," *Cognitive Psychology*, (12:3), 1980, pp. 306-355.

Gick, M.L. and Holyoak, K.J. "Schema induction and analogical transfer," *Cognitive Psychology*, (15:1), 1983, pp. 1-39.

Goldstone, R.L. and Sakamoto, Y. "The transfer of abstract principles governing complex adaptive systems," *Cognitive Psychology*, (46:4), 2003, pp. 414-466.

Goodhue, D.L., and Thompson, R.L. "Task-technology fit and individual performance," *MIS Quarterly.* (19:2), 1995, pp. 13-236.

Gordon, K., "Group Judgments in the Field of Lifted Weights," *Psychological Review*, (28:6), 1924, pp. 398-424.

Gurnee, H. "Maze Learning in the Collective Situation," *The Journal of Psychology*, (3), 1937, pp. 437-443.

Hamming, R.W. "You and Your Research," *Bell Communication Research Colloquium Seminar*, 1986.

Harary, F. *Graph Theory*, Addison Wesley, Reading, MA, 1994.

Howe, J. "The Rise of Crowdsourcing,", *Wired Magazine*, (14:6), 2006, pp. 1-4.

Holyoak, K.J. and Thagard, P. "Analogical Mapping by Constraint Satisfaction," in *Cognitive Modeling*, Polk, T.A. and Seifert, C.M. (Eds.), MIT Press, Cambridge, MA, 2002, pp. 849-909.

Holyoak, K.J. and Koh, K. "Surface and structural similarity in analogical transfer," *Memory and Cognition*, (15:4), 1987, pp. 323-340.

Hu, Y. F., "Efficient, High-Quality Force-Directed Graph Drawing," *The Mathematica Journal*, (10:1), 2006, pp.37-71.

Isler, J.M. "Career Chaos: Mapping Random Employment Trajectories in High Tech Work," *Paper presented at the annual meeting of the American Sociological Association,* NY, NY, August 2007, http://www.allacademic.com/meta/p183419_index.html.

Kazman, R., Asundi, J., and Klein, M. "Quantifying the costs and benefits of architectural decisions," in *Proceedings of the 23rd international Conference on Software Engineering*, IEEE Computer Society, Washington, DC, May 2001, pp. 297-306.

Kittur, A., Chi, E., Pendleton, B.A., Suh, B., and Mytkowicz, T. "Power of the Few vs. Wisdom of the Crowd: Wikipedia and the Rise of the Bourgeoisie," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, San Jose, California, 2007, pp. 453-462.

Knight, H. C. "A Comparison of the Reliability of Group and Individual Judgments," *Master's Thesis*, *Columbia University*, 1921.

Larkin, J.H., McDermott, J., Simon, D.P, and Simon, H.A. "Expert and Novice Performance in Solving Physics Problems," *Science* (208: 4450), 1980, pp. 1335-1342.

Lorge, I, Fox, D., Davitz, J., and Brenner, M. "A Survey of Studies Contrasting the Quality of Group Performance and Individual Performance, 1920-1957," *Psychological Bulletin*, (55:6), 1958, pp 337-372.

Lupyan, G, Rakison, D, H. and McClelland, J. L., "Language is not Just for Talking: Redundant Labels Facilitate Learning of Novel Categories," *Psychological Science*, (18:12), 2007, pp. 1077-1083.

Malone, T.W., Crowston, K., and Herman, G.A. *Organizing Business Knowledge: The MIT Process Handbook*, MIT Press, Cambridge, MA, 2003.

Nickerson, J.V. "Teaching the Integration of Information Systems Technologies," *IEEE Transactions on Education,* (49:2), May 2006, pp. 271-277.

Nickerson, J.V., Corter, J.E., Tversky, B., Zahner, D., and Rho, Y. "The Spatial Nature of Thought: Understanding Information Systems Design Through Diagrams," in *Proceedings of the 29th International Conference on Information Systems*, Boland, R., Limayem, M., and Pentland B. (Eds.), Paris, France, December 2008,

Novick, L. "Representational Transfer in Problem Solving," *Psychological Science*, (1:2), 1990, 128-132.

Novick, L. and Hmelo, C. "Transferring Symbolic Representations Across Nonisomorphic Problems," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, (20:6), 1994, pp. 1296-1321.

Ross, B.H. and Kennedy, P.T. "Generalizing From the Use of Earlier Examples in Problem Solving," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, (16:1), 1990, pp. 42-55.

Shank, R.C. *Dynamic Memory,* Cambridge University Press, New York, NY, 1982.

Schön, D.A. The Reflective Practitioner: How Professionals Think in Action, Basic Books, Jackson, TN, 1983.

Shanteau, J. "Why Do Experts Disagree?," in *Risk Behavior and Risk Management in Business Life*, Green, B., Cressy, R., Delmar, F., Eisenbert, T., Howcroft, B., Lewis, M., Schoenmaker, D., and Shanteau, J. (Eds.), Kluwer Academic Press, Dordrecht, The Netherlands, 2000, pp. 186-196.

Simon, D.P. and Simon, H.A. "Individual Differences in Solving Physics Problems," in *Children's Thinking: What Develops?*, Siegler, R.S. (Ed.), Lawrence Erlbaum Associates, Hillsdale, NJ, 1978.

Simon, H. A., and Chase, W G. "Skill in chess," *American Scientist*, (61:4), 1973, pp. 394-403.

Smith, L.B., Jones, S.S., Landau, B., Gershkoff-Stowe, L., and Samuelson, L. "Object Name Learning Provides On-the-Job Training for Attention," *Psychological Science*, (13:1), 2002, pp. 13-19.

Smith, D.E. and Nair, R. "The Architecture of Virtual Machines," *Computer*, (38:5), 2005, pp. 32–38.

Surowiecki, J. *The Wisdom of Crowds*, Anchor Books, New York, NY, 2004.

Sweller, J. "Cognitive Load During Problem Solving: Effects on Learning," *Cognitive Science,* (12:2), 1988, pp. 258-285.

Tversky, B., Zacks, J.M., Morrison, J.B., and Hard, B.M. "Talking About Events," in *Event Representation*, Pederson, E., Bohnemeyer, J., and Tomlin, R. (Eds.), Cambridge University Press, Cambridge, in press.

von Ahn, L. "Games with a Purpose," *Computer* (39:6), 2006, pp. 92-94.

Ullman, J.D. *Principles of Database and Knowledge-Base Systems, Vol. I*, Computer Science Press, Washington, DC, 1988.

Wagner, C. and Back, A. "Group Wisdom Support Systems: Aggregating the Insights of Many Through Information Technology," *Issues in Information Systems*, (9:2), 2008, pp. 343-350.

Wessel, D. "Prizes for Solutions to Problems Play Valuable Role in Innovation," *Wall Street Journal*, January 25, 2007.

Zigurs, I., and Buckland B.K. "A Theory of Task-Technology Fit and Group Support Systems Effectiveness," *MIS Quarterly*, (22:3), 1998, pp. 313-334.