MIS Quarterly

SPECIAL ISSUE: SOCIOMATERIALITY OF IS & ORGANIZING

TOWARD GENERALIZABLE SOCIOMATERIAL INQUIRY: A COMPUTATIONAL APPROACH FOR ZOOMING IN AND OUT OF SOCIOMATERIAL ROUTINES

James Gaskin
Brigham Young University, Provo, UT 84602 U.S.A. {james.gaskin@byu.edu}

Nicholas Berente
University of Georgia, Athens, GA 30602 U.S.A. {berente@uga.edu}

Kalle Lyytinen
Case Western Reserve University, Cleveland, OH 44106 U.S.A. {kalle@case.edu}

Youngjin Yoo
Temple University, Philadelphia, PA 19122 U.S.A. {youngjin.yoo@temple.edu}

Appendix A

Details of the Sociomaterial Lexical Elements for Design Routines

Activity Type

Activity type refers to the general purpose of each basic element within a sociomaterial routine. Activity types form an essential element of any part of a sociomaterial routine, because each different type of activity will typically invoke different combinations of social and material elements based on regularities and constraints that the activity type dictates while composing the other elements of that activity. For example, a brainstorming session may involve direct synchronous group communication, whereas validating a design may require only a single individual with a laptop. In order to render routines and their parts comparable across different contexts, we adopted a task classification scheme by McGrath (1984) that suggests four generic activity types: (1) generate, (2) choose, (3) negotiate, and (4) execute. Based on our field studies to analyze design routines, we have added to this list (5) validate (Bucciarelli 1994; Henderson 1991), (6) *transfer* (Daft 2009), and (7) *training* (Clark and Voogel 1985). As noted, an activity may include from one to several tasks.

Actor Configuration

Actor configuration refers to the number and grouping of the actors involved in the activity (Jin and Levitt 1996). The configuration of actors is not only a necessary social component of an activity, but also has important material regularities for how activities are composed. For example, collocated actors may use physical tools (e.g., whiteboards), whereas distributed users may use only digital tools (e.g., VOIP). Thus the social configuration of actors is inseparably connected with the material configuration of tools and related communication facilities. Actors can be exclusively configured as (1) an individual, (2) a group, (3) many individuals, (4) many groups, (5) or individuals and groups. The distinction between many individuals and a group is that a group shares a single functional purpose, whereas many individuals each have their own separate functional purpose. For example, many individuals can work on different parts of a product design, whereas a team can work on the task of reconciling or reviewing those parts.

Activity Location

Location refers to a locale where a sociomaterial activity takes place (Armstrong and Cole 2002). Like actor configuration, the location has both social and material implications and related regularities. Location often affects the tools used for manipulating work objects, and the number and types of actors involved. When actors are located in reference to each other, they it form an important factor for influencing how to coordinate the activity (Armstrong and Cole 2002; Eckert et al. 2005; Jonassen and Kwon 2001; Majchrzak et al. 2000; Perry and Sanderson 1998; Purvanova and Bono 2009; Robey and Jin 2004). Activities can be performed *collocated* (actors together in one place), *locally* (actors distributed at the host site), *remotely* (actors away from the host site), or in some combination of these—*mixed*.

Tool Modality

Tool modality refers to the underlying materiality of the functional affordances that the tool offers while being used in a sociomaterial activity. Tools come with specific *material features* (Markus and Silver 2008) that are afforded through their physical or digital modalities (Leonardi 2010; Leonardi and Barley 2008, 2010). Thus, for each tool, we code whether the modality of the tool is *digital* (the functional affordances of the tool are enabled via digital means—basically software) or *physical* (the functional affordances of the tool are enabled through physical means—like analogous hardware). In any given work context, the physical and the digital modality of work are tied together, and give meaning to each other (Davenport and Pearlson 1998; Jin 2002; Rennecker 2002; Robey and Jin 2004). In this vein, Robey et al. (2003) note that work capabilities can be enhanced by increasingly intertwining or differently intertwining physical and digital modalities. We place emphasis here on the modality rather than tool composition, because, inevitably, what is digital cannot be completely independent from the physical. In the end all digital is ultimately founded on some physical components (Leonardi 2010; Leonardi and Barley 2008).

Tool Affordance

Affordances refer to the opportunities for use offered, provided, or furnished by the technology for the user during a specific task. As such, we want to capture relational information between actors who participate in an activity, and what they do with the tools, by encoding tool affordances. Fundamentally, tools, whether digital or physical, can be used for multiple purposes by different actors within different tasks. Hence, tool affordance encodes a socially constructed (and interpreted through double-hermeneutic) relationship between the tool (technology) and a (set of) users in the studied activity (Gibson 1977; Markus and Silver 2008). Affordance is, therefore, not inherent to a tool, a task, or an actor, but to a context where the tool is being used by actors for a specific task (Carlo et al. 2011). At the same time, information concerning the modality of a tool is incomplete without information regarding the nature of its enactment. We integrated the list from Leonardi and Barley (2008) and Henderson and Cooprider (1994) while developing a list of tools affordances. We chose to base our list primarily on Henderson and Cooprider as they offer higher accuracy for describing alternative ways in which tools can be enacted by actors. Their list also overlaps considerably with Leonardi and Barley's typology and covers all of their affordances except a "storage" affordance, which we have added. We removed their "support" and "infrastructure" affordances, as we observed in our coding that they were subsumed by the "cooperative" use affordance. Thus we currently assume six affordances for tools: (1) representation, (2) analysis, (3) transformation, (4) control, (5) cooperative, and (6) storage.

Artifact Type

Artifact type refers to the modality of an artifact that is being manipulated as an object of work within a sociomaterial routine. During routine enactment, actors manipulate a variety of physical and digital artifacts, each intended to represent or support some aspect of their work. Again, artifact type is socially constructed and the artifacts can be primarily distinguished by their uses and the purposes they serve in different practices, such as communications (e-mail, memos, etc.), sketches, diagrams, requirements, and specifications (Rosenman and Gero 1996). In our taxonomy we employ five different artifact types: specification, draft/prototype, implementation, process planning, and knowledge. We capture information on artifact types because they are a central and often persistent component of sociomaterial routines. For example, documentation often carries over from one activity to another—being generated during one activity, updated during another, and used as input for yet another. These material artifacts can also have determining effects on social elements of activity. For example, process planning documents can dictate future communications and organizing. Thus the type of artifact being utilized or generated is intertwined with the social aspects of structuring an activity.

Dataflow

Finally, the dataflow refers to how a tool affects an artifact or relates to it. Artifacts can be an input for a tool, or an output of a tool, or they can be updated by the tool as part of the activity. We capture information on the tool/artifact connection, in order to detect the patterns of dataflow in a routine. Are artifacts being constantly updated? Used only a couple times as inputs? Or are they constantly being produced? While the tool/artifact connection is clearly a material relationship within a routine, it also has social ramifications. For example, the desired connection may affect which actors need to participate in an activity. If an activity needs to *output* a report of design possibilities, then this artifact may need to be generated by multiple actors, some with specific roles (such as architect, engineer, or designer), and the location of those actors will probably need to be collocated. Conversely, if a bill of materials needs to be generated (output), then a single individual (or automated procedure) can do this remotely.

References

Armstrong, D. J., and Cole, P. 2002. "Managing Distances and Differences in Geographically Distributed Work Groups," in *Distributed Work*, P. Hinds and S. Kiesler (eds.), Cambridge, MA: MIT Press, pp. 167-186.

Carlo, J. L., Lyytinen, K., and Boland, Jr., R. J. 2012. "Dialectics of Collective Minding: Contradictory Appropriations of Information Technology in a High-Risk Project," *MIS Quarterly* (36:4), pp. 1081-1108.

Clark, R. E., and Voogel, A. 1985. "Transfer of Training Principles for Instructional Design," *Educational Technology Research and Development* (33:2), pp. 113-123.

Daft, R. L. 2009. *Organization Theory and Design*, Mason, OH: South-Western Cengage Learning.

Davenport, T. H., and Pearlson, K. 1998. "Two Cheers for the Virtual Office," *Sloan Management Review* (39:4), pp. 51-65.

Eckert, C., Maier, A., and McMahon, C. 2005. "Communication in Design," in *Design Process Improvement*, J. Clarkson and C. Eckert (eds.), London: Springer-Verlag, 232-261.

Gibson, J. J. 1977. "The Theory of Affordances," Chapter 8 in *The Ecological Approach to Visual Perception,* Hillsdale, NJ: Lawrence Erlbaum Associates, 127-143.

Henderson, J. C., and Cooprider, J. 1984. "Dimensions of IS Planning and Design Aids: A Functional Model of CASE Technology," in *Information Technology and the Corporation of the 1990s: Research Studies*, T. Allen and M. S. Scott-Morton (eds.), Oxford, UK: Oxford University Press*, pp. 221-248.*

Henderson, K. 1991. "Flexible Sketches and Inflexible Data Bases: Visual Communication, Conscription Devices, and Boundary Objects in Design Engineering," *Science, Technology & Human Values* (16:4), pp. 448-473.

Jin, L. 2002. *An Interpretive Analysis of the Relationship Between Virtual and Material Representations in an E-Business Company: A Participant Observation Study*, unpublished Doctoral Dissertation, Georgia State University.

Jin, Y., and Levitt, R. E. 1996. "The Virtual Design Team: A Computational Model of Project Organizations," *Computational & Mathematical Organization Theory* (2:3), pp. 171-195.

Jonassen, D. H., and Kwon, H. 2001. "Communication Patterns in Computer Mediated Versus Face-to-Face Group Problem Solving," *Educational Technology Research and Development* (49:1), pp. 35-51.

Leonardi, P. M. 2010. "Digital Materiality? How Artifacts Without Matter, Matter," *First Monday* (15:6-7).

Leonardi, P., and Barley, S. 2008. "Materiality and Change: Challenges to Building Better Theory About Technology and Organizing," *Information and Organization* (18:3), pp. 159-176.

Leonardi, P. M., and Barley, S. R. 2010. "What's Under Construction Here? Social Action, Materiality, and Power in Constructivist Studies of Technology and Organizing," *The Academy of Management Annals* (4:1), pp. 1-51.

Liu, P., and Pentland, B. 2011. "Dynamic Capabilities and Business Processes: A Trajectory View," in *Proceedings of the 17th Americas Conference on Information Systems*, Detroit, MI, August 3-4.

Majchrzak, A., Rice, R. E., Malhotra, A., King, N., and Ba, S. 2000. "Technology Adaptation: The Case of a Computer-Supported Inter-Organizational Virtual Team," *MIS Quarterly* (24:4), pp. 569-600.

Markus, M. L., and Silver, M. S. 2008. "A Foundation for the Study of IT Effects: A New Look at DeSanctis and Poole's Concepts of Structural Features and Spirit," *Journal of the Association for Information Systems* (9:10), Paper 5.

McGrath, J. E. 1984. *Groups: Interaction and Performance*, Englewood Cliffs, NJ: Prentice-Hall.

Perry, M., and Sanderson, D. 1998. "Coordinating Joint Design Work: The Role of Communication and Artefacts," *Design Studies* (19:3), pp. 273-288.

Purvanova, R. K., and Bono, J. E. 2009. "Transformational Leadership in Context: Face-to-Face and Virtual Teams," *The Leadership Quarterly* (20:3), pp. 343-357.

Rennecker, J. A. 2002. "The Situated Nature of Virtual Teamwork: Understanding the Constitutive Role of 'Place' in the Enactment of Virtual Work Configuration," *SPROUTS: Working Papers on Information Environments, Systems and Organizations* (2:3), pp. 115-139.

Robey, D., and Jin, L. 2004. "Studying Virtual Work in Teams, Organizations, and Communities," *The Handbook for Information Systems Research*, Hershey, PA: IGI GLobal, pp. 150-165.
Robey, D., Schwaig, K. S., and Jin, L. 2003. "Intertwining Material and Virtual Work," *Information and Organization* (13:2), pp. 111-129.
Rosenman, M. A., and Gero, J. S. 1996. "Modelling Multiple Views of Design Objects in a Collaborative Environment," *Computer-Aided Design* (28:3), pp. 193-205.

Appendix B

A Step by Step Illustration of the Approach[1]

This appendix illustrates, in step-by-step fashion, how to analyze activity sequences and compositions when a representation of a routine, in terms of its sociomaterial elements, has been derived using the MetaEdit+ visualization tool. For this step by step guide, we will refer to the case of a design routine from the construction industry that is repeated for each construction project—and, thus, is considered routine. The data from which most[2] output and illustrations will be derived comes from a collection effort at a large American construction company (pseudonym: BuildMAX) over a five year period. We collected data about MEP (Mechanical, Electrical, and Plumbing) coordination design for four standard (but large) projects. MEP design is an integral part of construction design that is becoming increasingly virtually coordinated (prior to physical implementation). MEP coordination at BuildMAX is continuously evolving as the activities are refined, removed, or institutionalized. Thus, the virtual MEP coordination at BuildMAX offers an ideal opportunity to study variation and evolution of sociomaterial routines. The steps below are expanding upon the summary offered in Table 3 in the main text.

Step 1. Query MetaEdit+ Database

MetaEdit+ allows us to write custom scripts to query the databases of sociomaterial routines for all pertinent information relative to a given set of activities. We have written multiple scripts in MetaEdit+ query language to scrape relevant data from diagrams depicting routines (like the one shown in Figure 1 in the main text), including property values of activities, actors, tools, affordances, and artifacts associated with each activity. The scripts also maintain and estimate the correct sequence for the activities for further sequencing based on order information (i.e., what kind of activity precedes another). A generic and customizable script is provided in Appendix C. The output of the script will be a raw, tab-delimited dataset like the snippet of one shown in Table B1. Among other information, row 1 of Table B1 tells us that this was a transfer activity done remotely, during which the action was to store some prototype drawings (2D or 3D models) via ftp.

Table B1. Snippet of MetaEdit+ Output

1	Transfer	Remote	Post drawings	1 group	Storage	ftp	Digital	Output	Persistent	Prototype	2D or 3D models
2	Validate	Remote	Validate design model	many groups	Transform	Modeling Tool	Digital	Update	Persistent	Specification	Architectural and Construction Models
3	Generate	Remote	Generate more model info	many groups	Transform	Modeling Tool	Digital	Update	Persistent	Specification	Architectural and Construction Models

[1] The examples provided here build on prior work (Gaskin et al. 2010a; Gaskin et al. 2011a; Gaskin et al. 2010b; Gaskin et al. 2011b), but the analyses have been further refined for this manuscript and we used additional and new data.

[2] Figure B2 and Table B3 come from a separate data collection effort involving the sociomaterial routines of four large American companies.

Step 2a. Restructure Data into Sequences Using Excel

We then analyze and reformat the output of our scripts through a series of automated procedures written in VBA for Excel to render the data into a format that is analyzable during sequence analysis using Clustal sequencing tools (described below). By applying an interpretation key—essentially replacing the element values as shown in Table B1 with their respective lexical values—we can restructure the dataset into strings of sequenced lexical values. Automated procedures can be written in VBA to concatenate values based on their parent activity. Thus, if there are 25 activities in a given routine, then there will be 25 sequences once the script is finished running. The final format of the sequences appears as shown in Figure B1.

```
>s1_Post_drawings
TfrRmtGrpStoDigOutPro
>s2_Validate_design_model
ValRmtGpxTraDigUpdUpdProSpe
>s3_Generate_model_info
GenRmtGpxTraDigUpdUpdProSpe
```

Figure B1. Examples of Sequences in Clustal Format

Step 2b. Explore Descriptive Statistics Using Pivot Tables and Pivot Charts

Using programmed procedures in Excel, we can also restructure the data to be analyzable for descriptive statistics provided in pivot tables and pivot charts. Descriptive analyses reveal patterns in data such as the frequency of specific types of affordances for each activity type, or the percent of each type of modality for sociomaterial routines from different organizations (Figure B2). These types of analyses make the detection of patterns much easier. They are also particularly useful for comparing activities from the same project across multiple stages, to see how social and material aspects of activities evolve over time within an organization (Table B2). Obtaining this kind of information using traditional interpretivist methods would be incredibly time-consuming and prone to human error.

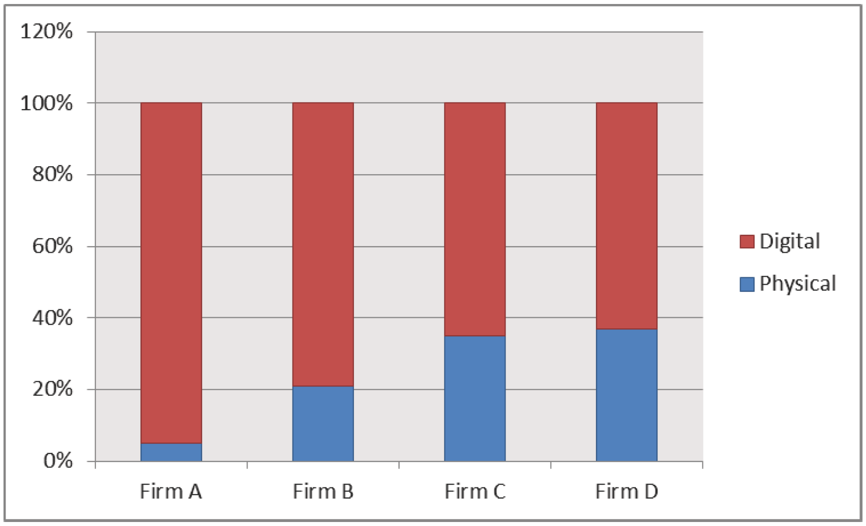


Figure B2. Cross Sectional Example of Descriptive Statistics Using Pivot Charts

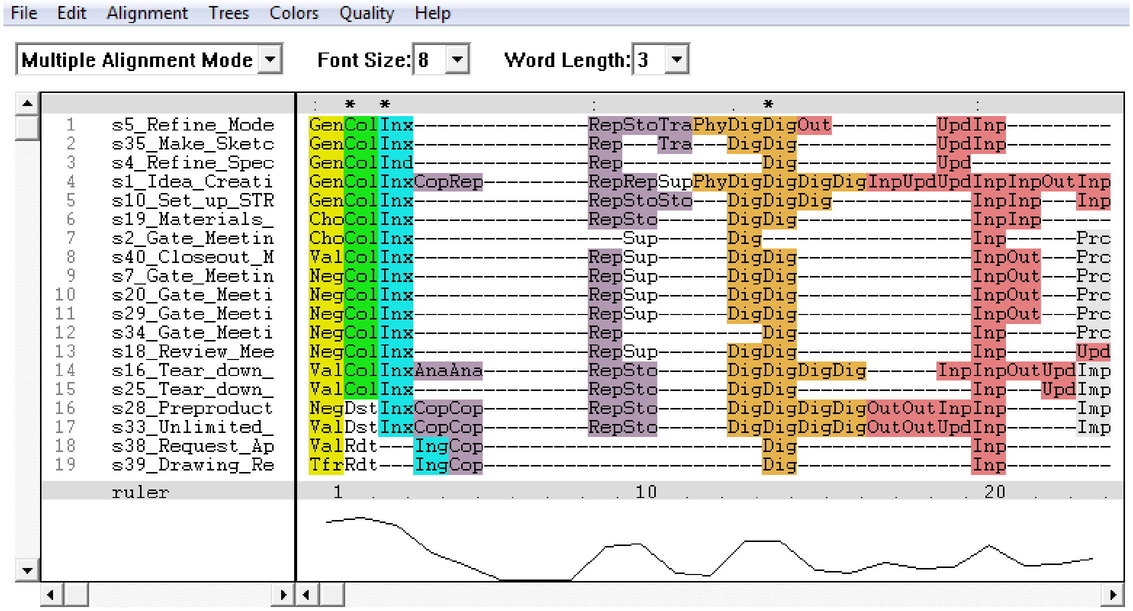
Table B2 provides a zoomed out perspective illustrating that the amount of work done by a single group was highly dominant in 2007 when BuildMAX was just adjusting to the new mode of virtual work. However, as they became more familiar and comfortable with the tasks and deliverables, they no longer had to meet together as often in a single body to work things out. They were able to modularize and subcontract the majority of their routines. This is evident in the sharp increase of work done by "many groups." With this insight now available, we can then zoom back in to gain a better situated understanding of the change in routines: "The subcontractors are now sending us models where they've already coordinated everything, so…we don't have to spend all our energy on [it], and so now, we can take all our time, and really focus on…things they're not paying attention to" (Designer 4/13/2011).

Table B2. Longitudinal Example of Descriptive Statistics Using Pivot Tables

Year	1 Group	1 Individual	Many Groups	Total
MEP 2007	90%	0%	10%	100%
MEP 2008	38%	0%	62%	100%
MEP 2009	20%	15%	65%	100%
MEP 2011	17%	4%	79%	100%

Step 3. Multiple Alignment in ClustalTXY

After conducting descriptive analyses, we import the data into ClustalTXY (Wilson 2001, 2006; Wilson et al. 2005) for multiple alignment and sequence analysis. ClustalTXY is a derivative of ClustalX and ClustalW (Larkin et al. 2007), both widely used biological sequence analysis tools that detect protein and nucleotide molecules and their structures (Aiyar 2000; Thompson et al. 2002).[3] ClustalTXY (henceforth Clustal) has been expanded from ClustalX and ClustalW in order to enable the analysis of multiple types of sequences in social sciences. For example, Wilson (2001) used Clustal to analyze and cluster day-to-day activity patterns among Canadian women. The reliability of sequence analyses using Clustal has been demonstrated by Wilson (2006). Clustal performs a pairwise alignment of the sequences in order to construct a similarity matrix that is then converted into distance scores. It compiles multiple alignments based on the branching pattern of a tree calculated from pairwise distances and other conventions that affect alignment scores (Wilson 2001), including a user specified weight matrix.[4] For instructions on using Clustal, refer to Wilson et al. (2005). Figure B3 is a screenshot of the user interface after multiple alignment has been executed. The colors represent different taxonomical elements. The three-character abbreviations represent the specific values of the taxonomic elements presented in Table 2 of the article. The dashes represent gaps inserted in order to arrive at an optimal matching solution. The graph on the bottom indicates the level of alignment.



Multiple Alignment Mode ▾ Font Size: 8 ▾ Word Length: 3 ▾

```
                           :  *  *                 :        .  *                 :
 1  s5_Refine_Mode   GenColInx---------------RepStoTraPhyDigDigOut---------UpdInp---------
 2  s35_Make_Sketc   GenColInx---------------Rep---Tra---DigDig------------UpdInp---------
 3  s4_Refine_Spec   GenColInd---------------Rep------------Dig------------Upd------------
 4  s1_Idea_Creati   GenColInxCopRep---------RepRepSupPhyDigDigDigDigInpUpdUpdInpInpOutInp
 5  s10_Set_up_STR   GenColInx---------------RepStoSto---DigDigDig---------------InpInp---Inp
 6  s19_Materials_   ChoColInx---------------RepSto------DigDig---------------InpInp------
 7  s2_Gate_Meetin   ChoColInx------------------Sup------Dig------------------Inp------Prc
 8  s40_Closeout_M   ValColInx---------------RepSup------DigDig---------------InpOut---Prc
 9  s7_Gate_Meetin   NegColInx---------------RepSup------DigDig---------------InpOut---Prc
10  s20_Gate_Meeti   NegColInx---------------RepSup------DigDig---------------InpOut---Prc
11  s29_Gate_Meeti   NegColInx---------------RepSup------DigDig---------------InpOut---Prc
12  s34_Gate_Meeti   NegColInx---------------Rep------------Dig---------------Inp------Prc
13  s18_Review_Mee   NegColInx---------------RepSup------DigDig---------------Inp------Upd
14  s16_Tear_down_   ValColInxAnaAna---------RepSto------DigDigDigDig------InpInpOutUpdImp
15  s25_Tear_down_   ValColInx---------------RepSto------DigDig---------------Inp---UpdImp
16  s28_Preproduct   NegDstInxCopCop---------RepSto------DigDigDigDigOutOutInpInp------Imp
17  s33_Unlimited_   ValDstInxCopCop---------RepSto------DigDigDigDigOutOutUpdInp------Imp
18  s38_Request_Ap   ValRdt---IngCop------------------------Dig---------------Inp---------
19  s39_Drawing_Re   TfrRdt---IngCop------------------------Dig---------------Inp---------
    ruler            1 . . . . . . . . 10 . . . . . . . . . 20 . . .
```

Figure B3. Clustal User Interface (After Multiple Alignment)

[3] ClustalTXY is free and open source, downloadable from the following ftp address: ftp://ftp-igbmc.u-strasbg.fr/pub/julie/clustaltxy.zip.

[4] The weight matrix is used to specify scores for aligning two values. Weights specified in our matrix include: 0 for aligning values from different elements, 5 for aligning different value from the same element, and 10 for aligning identical values. However, the results were robust to using/not using the weight matrix.

Step 4. Analysis of Variation

Clustal produces the alignments between each pair of sequences for the entire set of sequences. If the sequences are identical, the alignment percent will equal 100; if they have absolutely no common elements, the alignment percent will be 0 (anything in between will depend on your weight matrix). This large set of pairwise permutations of alignments can then be summarized by groups in Excel to calculate the average percent alignment between groups of sequences. For example, if we align sequences in Clustal from four different organizations, we can see what is the average alignment of sequences within an organization and across multiple organizations. These average alignments speak to the degree of variation exhibited in sociomaterial activities. Such an analysis is offered in Table B3. For example, this table shows that, on average, the activities from routines at Firm A are 47 percent different from the ones at Firm B. Biologists use this technique to describe the percent difference (or similarity) between species. For example, a human and a chimpanzee have an average of about 97 percent of their DNA in common; whereas a human and a chicken only have about 60 percent in common. We can also concatenate the entire routine (e.g., all seven-tuples for all activities for MEP 2007) and align that with another entire routine. This is useful for comparing changes in the same type of routine over time (i.e., evolution). Using the MEP example, we can observe the rate of evolution between MEP routines from year to year in Figure B4. In this example, the rate of evolution is declining as the MEP team refines and institutionalizes their practices. These percentages represent the degree of difference between the sequential configurations of elements for each activity from year to year. Put in more practical terms, it represents (at a very detailed level) the extent to which this year's project differs from last year's.

Table B3. Cross-Company Analysis of Variation

	Firm A	Firm B	Firm C	Firm D
Firm A	34%			
Firm B	47%	33%		
Firm C	44%	48%	4%	
Firm D	46%	48%	43%	31%

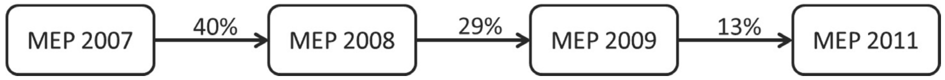


Figure B4. Rate of Evolution in the MEP Process at BuildMAX

Step 5. Sequence Analysis Using Transition Matrices

In addition to the sequence alignment and the descriptive statistics, we can also produce Markov chains, or transition matrices, to describe the state changes in activity types, actor configurations, or locations.[5] Transition matrices estimate the probability of one state following another, and are thus the truest *sequence* analysis. For example, transition matrices can answer the questions

- For BuildMAX, what type of activity is most likely to follow a "validate" activity?
- How often do validate activities follow generate activities?

The transition matrix takes as an input the temporal sequence of all activity types (or whichever element you desire), and then estimates the probability that one state will follow another. Building a transition matrix by hand is not feasible for even moderately sized processes (anything greater than about 20 activities). Accordingly, we have written an automated procedure in Excel that generates the frequency transition matrix, expected transition matrix, and probability transition matrix. Tables B4-6 represent each of these three kinds of matrices,[6] while Figure B5 shows one way to visually depict a probability transition matrix.

[5] These are the elements that occur only once per activity, and thus their precise order can be controlled, whereas, for the other elements, they may occur multiple times in a given activity, and not in any particular temporal order. Thus, these other elements are not appropriate for state change analysis that is completely dependent upon temporal order.

[6] Not all activity types are represented in these transition matrices because not all activity types were used during this particular routine.

The *frequency* transition matrix is the exact count of state changes from one activity type to another. The probability transition matrix is the likelihood (percent) that one activity type will follow another based on the frequency transition matrix. The *expected* transition matrix is the expected transition frequency between each pair of activity types based solely on the count of each activity type in the routine. For example, if we know that a routine with 10 activities has only two types of activities—*choose* and *negotiate*—both occurring five times, then we can expect *choose* to follow *negotiate* 50 percent of the time. However, if *choose* occurs eight times and *negotiate* occurs two times, then we would only expect *choose* to follow *negotiate* 20 percent of the time. When there are more than two values, the calculations get more difficult; hence, we have written procedures to automate these calculations. Additionally, using the frequency transition matrix (what the state changes actually were) and the expected transition matrix (what we would expect based on the number of each value), we can then conduct a chi-square difference test to determine if there is interdependence and predictability among the activities or whether the order is essentially random.

We can also take observed transition matrices from multiple processes and compare them to determine if the sequence of activities is roughly equivalent or statistically different. Using this approach we can actually produce p-values to determine structural (and sequential) equivalence. This is done using software called UCINet[7] which allows quadratic correlation tests between matrices. For example, MEP 2009 and 2011 are not statistically similar in their sequential transitions (Pearson Correlation = 0.35, p = 0.25), whereas MEP 2007 and 2011 are significantly dissimilar (Pearson Correlation = -0.50, p = 0.04), as shown in Table B7. In an ongoing research effort, we are currently exploring the possibility and usefulness of using transition matrices to simulate sociomaterial routines (Liu and Pentland 2011). Such simulations would open up possibilities to answer new questions about the effects of inserting into or removing specific social or material aspects from a process.

Table B4. Frequency Transition Matrix

	Generate	Choose	Validate	Negotiate	Transfer
Generate	7	1	8	0	0
Choose	1	0	0	1	0
Validate	3	0	6	4	1
Negotiate	4	1	0	1	0
Transfer	0	0	1	0	0

Table B5. Probability Transition Matrix

	Generate	Choose	Validate	Negotiate	Transfer
Generate	44%	6%	50%	0%	0%
Choose	50%	0%	0%	50%	0%
Validate	21%	0%	43%	29%	7%
Negotiate	67%	17%	0%	17%	0%
Transfer	0%	0%	100%	0%	0%

Table B6. Expected Transition Matrix

	Generate	Choose	Validate	Negotiate	Transfer
Generate	6.2	0.8	6.2	2.5	0.4
Choose	0.8	0.1	0.8	0.3	0.1
Validate	5.4	0.7	5.4	2.2	0.4
Negotiate	2.3	0.3	2.3	0.9	0.2
Transfer	0.4	0.1	0.4	0.2	0.0

[7] http://www.analytictech.com/ucinet/.

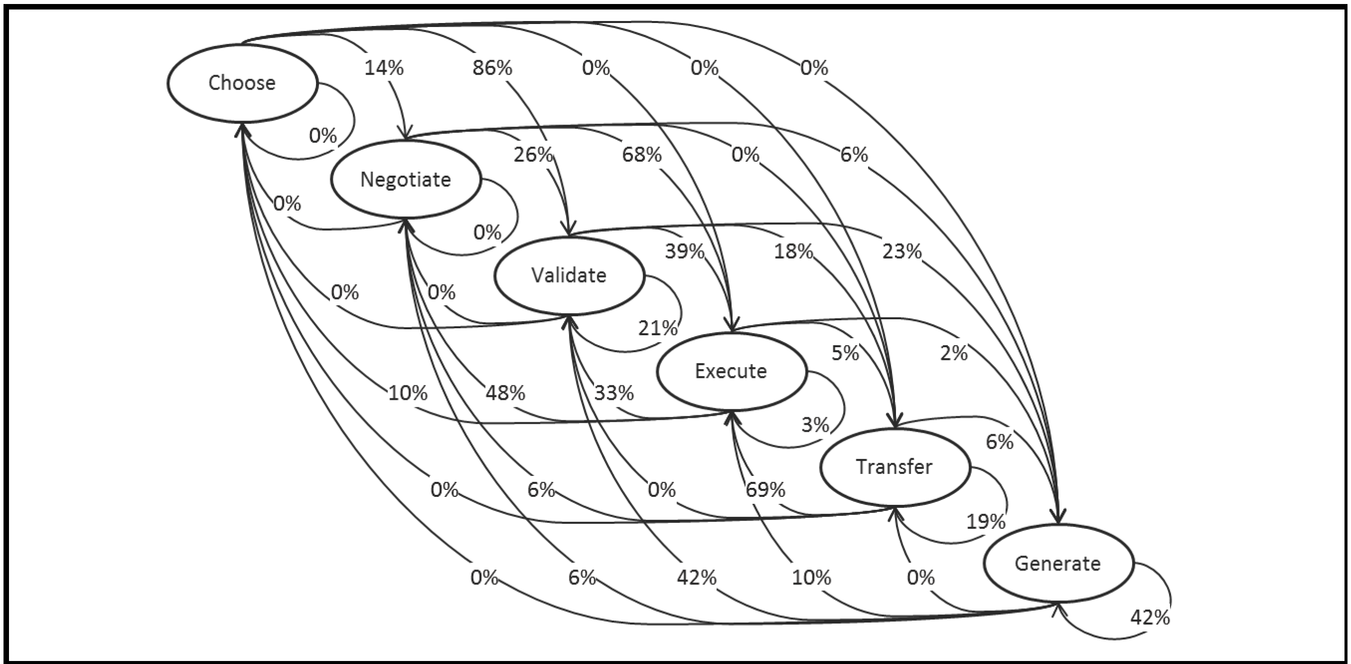


Figure B5. Activity Transition Matrix (Markov Chain)

Table B7. Pearson Correlations (and p-values) from QAP Analysis in UCINet

	MEP2007	MEP2008	MEP2009
MEP2008	0.16 (0.38)		
MEP2009	-0.47 (0.17)	0.32 (0.29)	
MEP2011	-0.50 (0.04)	0.04 (0.36)	0.35 (0.25)

Step 6. Cluster Analysis in MEGA

Finally, Clustal produces a dendrogram file that can be used to draw hierarchical and phylogenetic trees in DNA mapping programs like MEGA (Kumar et al. 2004).[8] These trees offer insights into the clustering structure of sequences of activities in routines. In practical terms, this helps identify activities that utilize the same resources and have the same intent. Such information may be useful to project managers seeking to modularize to reduce design costs. Notice in Figure B6 that routine clusters are driven primarily by activity type. For example, the clusters in the six o'clock position are all transfer activities, but are from four different projects from four different years. These trees also show degrees of difference between sequences. Branch lengths are an estimate of the distance (or difference) between the sequence on that branch and the nearest parent branch. For example, the MEP 2009 and 2011 validate activities (eleven o'clock) are quite different from the others, as is the sequence labeled "s2 2007 Validate" (around the two o'clock position), which appears on its own.

To test the reliability of these clusters, we may resample the sequences using a bootstrap feature available in Clustal. This feature collapses all bootstrapped trees to form a consensus tree—a tree that represents consistent "true" structure. The output of the procedure is a PHYLYP tree file (.phb), which can be viewed in MEGA. The consistency of cluster structures is represented by the number of times a particular cluster appeared divided by the total number of resamples. These percentages are nonparametric confidence values for the reliability of each cluster—much like Cronbach's alpha, or composite reliability scores. Accordingly, percentages above 0.70 indicate reliable clusters.[9] Finally, trees should be visualized as "unrooted," because this part of the approach is intended for detecting cluster structures, whereas rooted trees imply ancestral relations, or a common ancestor.

[8] http://www.megasoftware.net/.

[9] In Phylogenetics, 80% is the standard threshold.

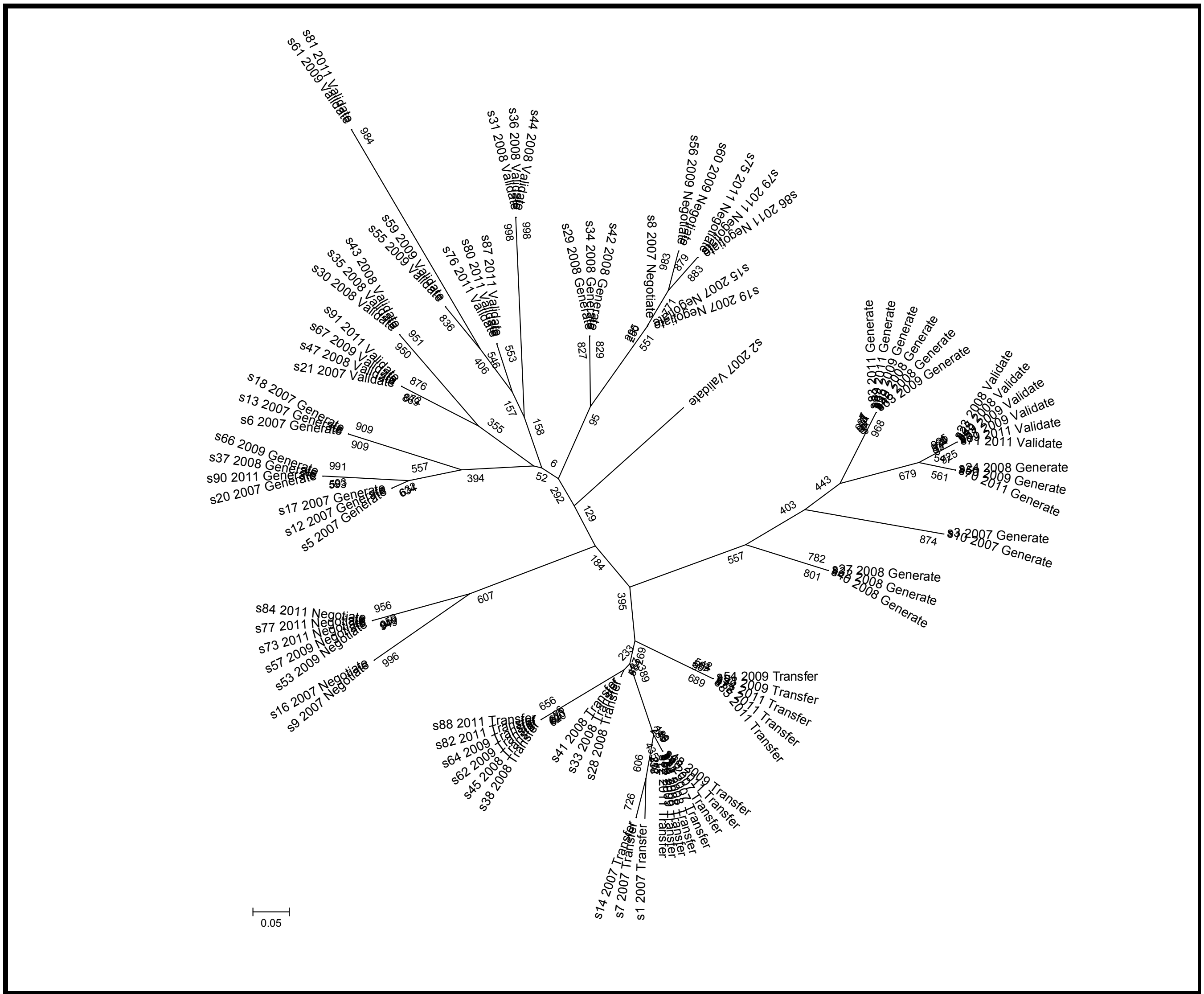


Figure B6. Unrooted Consensus Tree (Numbers Represent Presence out of 1000 Resamples)

References

Aiyar, A. 2000. "The Use of CLUSTAL W and CLUSTAL X for Multiple Sequence Alignment," *Methods in Molecular Biology* (132), pp. 221-241.

Bucciarelli, L. L. 1994. *Designing Engineers*, Cambridge, MA: MIT Press.

Gaskin, J., Lyytinen, K., Thummadi, V., Schutz, D., Yoo, Y., Weiss, A., and Berente, N. 2010a. "Design DNA Sequencing: A Set of Methodological Artifacts for Sequencing Socio-Technical Design Routines," in *Proceedings of the 31st International Conference on Information Systems*, St. Louis, MO, December 12-15.

Gaskin, J., Lyytinen, K., Yoo, Y., Shiv, O., and Zhang, Z. 2011a. "Evolution of Digitally-Enabled Design Processes: The Case of a Large Design and Construction Firm," paper presented at the 71st Annual Academy of Management Meeting, San Antonio, TX, August 12-16.

Gaskin, J., Schutz, D., Berente, N., Lyytinen, K., and Yoo, Y. 2010b. "The DNA of Design Work: Physical and Digital Materiality in Project-Based Design Organizations," in *Academy of Management Best Paper Proceedings*, Montreal, Canada, August 6-10.

Gaskin, J., Thummadi, B. V., Lyytinen, K., and Yoo, Y. 2011b. "Digital Technology and the Variation in Design Routines: A Sequence Analysis of Four Design Processes," in *Proceedings of the 32nd International Conference on Information Systems*, Shanghai, China, December 4-7.

Kumar, S., Tamura, K., and Nei, M. 2004. "MEGA3: Integrated Software for Molecular Evolutionary Genetics Analysis and Sequence Alignment," *Briefings in Bioinformatics* (5:2), pp. 150-163.

Larkin, M., Blackshields, G., Brown, N., Chenna, R., McGettigan, P., McWilliam, H., Valentin, F., Wallace, I., Wilm, A., and Lopez, R. 2007. "Clustal W and Clustal X Version 2.0," *Bioinformatics* (23:21), pp. 2947-2948.

Thompson, J. D., Gibson, T. J., and Higgins, D. G. 2002. *Multiple Sequence Alignment Using ClustalW and ClustalX*, Hoboken, NJ: John Wiley & Sons, Inc.

Wilson, C. 2001. "Activity Patterns of Canadian Women: Application of ClustalG Sequence Alignment Software," *Transportation Research Record: Journal of the Transportation Research Board* (1777), pp. 55-67.

Wilson, C. 2006. "Reliability of Sequence-Alignment Analysis of Social Processes: Monte Carlo Tests of ClustalG Software," *Environment and Planning A* (38:1), pp. 187-204.

Wilson, C., Harvey, A., and Thompson, J. 2005. "Clustalg: Software for Analysis of Activities and Sequential Events," paper presented at the Workshop on Sequence Alignment Methods, Halifax, Nova Scotia, October.

Appendix C

Generic Code for Scraping MetaEdit+ Database

```
Report ‘allProperties’
foreach.()
{ newline
newline ‘------------------‘ newline
 ‘’‘’ id ‘’‘ <‘ type ‘>’
 newline
 do :()
 { if id;1 <> id then
 ‘ “’ type ‘”:  ‘ id
 newline
 endif
 }
       if not ~() then
              ‘ takes part in no relationships’
              newline
       else
              ‘ takes part in relationship(s):’
              newline
              dowhile ~()
              {      ‘      ’
                     do >()
                     {      if id then ‘’‘’ id ‘’‘ ’ endif
                            ‘<‘ type ‘>’
                     }
                     ‘ with object(s) ‘
                     dowhile ~().()
                     {      if id then ‘’‘’ id ‘’‘ ’ endif
                            ‘<‘
                            type
                            if ‘always true’ then ‘>’ endif
                            ‘, ‘
                     }
       if not ~() then
              ‘ takes part in no relationships’
              newline
       else
              ‘ takes part in relationship(s):’
              newline
              dowhile ~()
              {      ‘      ’
                     do >()
                     {      if id then ‘’‘’ id ‘’‘ ’ endif
                            ‘<‘ type ‘>’
                     }
                     ‘ with object(s) ‘
                     dowhile ~().()
                     {      if id then ‘’‘’ id ‘’‘ ’ endif
                            ‘<‘
                            type
                            if ‘always true’ then ‘>’ endif
                            ‘, ‘
                     }
                     newline
              }
endif
newline
              }
endif
}
endreport
```