Carnegie Mellon

Construct – A Multi-agent network model for the co-evolution of agents and socio-cultural environments*

CASOS Technical Report

Craig Schreiber, Siddhartha Singh and Kathleen M. Carley

May 2004

CMU-ISRI-04-109

Carnegie Mellon University
School of Computer Science
ISRI – Institute for Software Research International
CASOS – Center for Computational Analysis of Social and Organizational Systems

Abstract

The Construct technical report describes the Construct model and lists the theories which it incorporates. Scientific literature that has used the model is listed as well as representative examples of real-world use within organizations. The report also defines the input and output variables and describes the various input and output files used with Construct. System requirements and performance characteristics are provided. Illustrative examples furnish a mini-tutorial on how to run virtual experiments with Construct.

* The original version of Construct was developed under NSF IST-8607303. More recently, research to extend and validate Construct has been supported in part by the National Science Foundation NSF IRI9633 662, 662, NSF KDI IIS-9980109, the NSF IGERT program in CASOS, and by the Office of Naval Research (ONR), United States Navy Grant No. N00014-97-1-0037, Army Research Lab Contract No. DASW01-00-K-0018 and NASA. Additional support was provided by CASOS and ISRI at Carnegie Mellon University. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Office of Naval Research, the Army Research Labs, NASA, the National Science Foundation, or the U.S. government.

Keywords: organization theory, computational analysis, social networks, multi-agent simulation, socio-cultural environment, information diffusion, socio-technical environment, model validation

Table of Contents

1. Model Description 2
1.1 Agents 2
1.2 Groups and size 2
1.3 Interaction mechanism 2
1.4 Knowledge 3
1.5 Task 3
1.6 Validation 4
1.7 Literature 4
1.8 Real-World Use 4
2. Model Input Variables and Output Files 4
2.1 Input Variables 4
2.2 Output Files 11
3. System Requirements 15
4. Performance Characteristics 15
5. Illustrative Examples 16
5.1 Model random generation 16
5.2 Create a knowledge base for input 17

References 19
Appendix A – Output file detail 21

1. Model Description

Construct is a multi-agent network model for the co-evolution of agents and socio-cultural environments. Construct is designed to capture dynamic behaviors in organizations with different cultural and technological configurations. Construct models groups and organizations as complex systems and captures the variability in human, technological and organizational factors through heterogeneity in information processing capabilities, knowledge and resources. The non-linearity of the model generates complex temporal behavior due to dynamic relationships among agents. Agents are decision-making units and can represent various levels of analysis such as individuals, groups or organizations. The Construct model is grounded in structuration theory (Giddens, 1984), social information processing theory (Salancik and Pfeffer, 1978) and symbolic interactionism (Manis and Meltzer, 1978).

1.1 Agents

Agents are characterized by their information processing capabilities (see Table 1).

Table 1: Agents by information processing capabilities

Agent	Initiate	Send	Sending Availability	Receive (Learn)	Receiving Availability
Humans	Yes	Yes	1:1	Yes	1:1
Book	No	Yes	1:N	No	n/a
Flyer (5 fact limit)	No	Yes	1:N	No	n/a
Avatar	Yes	Yes	1:N	No	n/a
Database	No	Yes	1:N	Yes	1:1

Communication style in the interface is chosen according to sending availability. Anytime there is a technology agent present then 1:N is chosen, otherwise 1:1.

1.2 Groups and size

Currently the maximum number of groups is limited to 3 and maximum number of agents is limited to 101.

1.3 Interaction mechanism

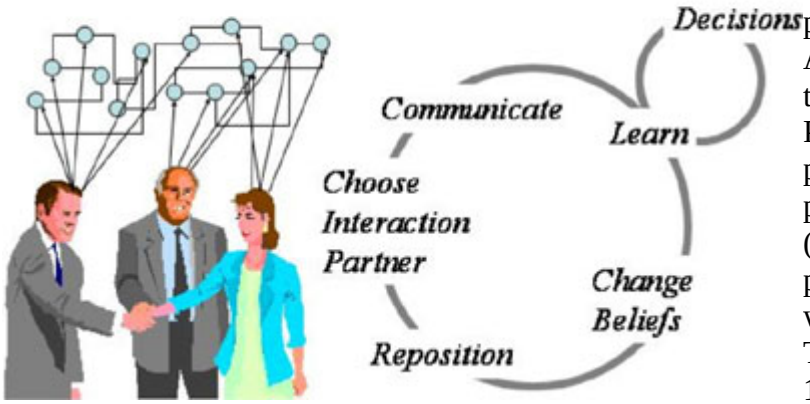
Construct models interaction based on constructuralism (Carley, 1986a; Carley, 1986b; Carley, 1990, Carley, 1991b, Carley, 1995). Constructuralism states that the socio-cultural environment is continually being constructed and reconstructed through individual cycles of action, adaptation and motivation, see Figure 1. Agent motivation is represented by the choice of interaction partners. Agents take action by interacting and exchanging knowledge. Agents adapt by learning, changing beliefs and repositioning themselves in the social network. As these cycles are continually repeated, change occurs and social structure and culture co-evolve. The process of change in Construct is such that agent interaction leads to environmental change which in turn effect subsequent agent interaction.

The basic interaction mechanism embodies three empirical generalizations: (1) knowledge acquisition occurs through interaction (Festinger, 1950; Granovetter, 1974), (2) homophily

(Lazarsfeld and Merton, 1954; McPherson and Smith-Lovin, 1987) and (3) social relativity (Merton, 1968; Festinger, 1954).

Information seeking is incorporated into Construct as an interaction choice opposite to homophily. An agent that is information seeking will choose to interact with other agents who are dissimilar to the information seeker.

Figure 1: Construct Interaction Cycle



Five moderators of the interaction mechanism are available: forgetting, proximity, transactive memory, referrals and access. Forgetting affects the adaptation phase by moderating learning. Agents can forget information they previously learned. Proximity affects the motivations phase. Research has shown that proximity moderates interaction (Festinger, et al., 1950). A proximity moderator can be set for within and between groups. Transactive memory (Wegner, 1986) affects the motivation phase. Each agent has the ability to know and learn about other agents in the environment. The agents can understand who knows what or who has what resource, etc. Agents in the model can use their transactive memory to inform their choice of interaction partner. In addition, research has provided support that transactive memory impacts group performance (Moreland, et al., 1996). Referrals affect the motivation phase. Referrals have been shown to be a source for retrieving information (Constant, et al., 1996). When agents receive referrals, either from another agent or a technology, they subsequently choose the agent being referred to as the interaction partner in the next time period. Access affects the motivation phase. Access sets a strict probability for an agent's ability to choose another agent. In this way, asymmetries in agent accessibility can be modeled.

1.4 Knowledge

Knowledge is represented in a binary string. Each agent knows a percentage of the true knowledge bit string. As the agents interact they exchange the knowledge they have. It is the possession of knowledge that determines their decision in the task. Agents can also be assigned a perception of whether a particular bit has a positive, neutral or negative bearing on the task decision. Currently knowledge is limited to 500 facts.

1.5 Task

Agents perform the binary classification task. Organizational researchers have widely studied the binary classification task (Hollenbeck, Ilgen, Tuttle and Sego, 1995). This task is an unbiased decomposition task whereas agents decide if there are more 1's or 0's in a binary string.

Up to 25 binary classification tasks can be defined for a particular knowledge bit string. Specific bits in the knowledge bit string can be identified as being relevant to a task.

1.6 Validation

The Construct-TM model has been scientifically validated several times (Carley, 1990; Carley and Krackhardt, 1996; Carley and Hill, 2001; Schreiber and Carley, in press). The first validation (Carley, 1990) used Kapferer's Zambia tailor shop data. By combining individual and social considerations, the model was able to predict observed changes in human interactions. The latest validation (Schreiber and Carley, in press) found significant correlation between communication patterns within real-world organizations and agent interactions within the model.

1.7 Literature

This is a sample of published works that used Construct.

Carley (1990; 1991a; 1991b; 1995; 1996; 1999; 2002)
Carley and Hill (2001)
Carley and Krackhardt (1996)
Carley and Schreiber (2002)
Kaufer and Carley (1993) – the most extended use of Construct
Schreiber and Carley (in press; 2003)

1.8 Real-World Use

The model has been used for analysis and consulting in industry (health care, aerospace, consulting, professional associations, financial), non-profit and emergency response (charity foundation, American Red Cross), higher education (universities), military (DARPA, ONR) and government (NSF, NASA).

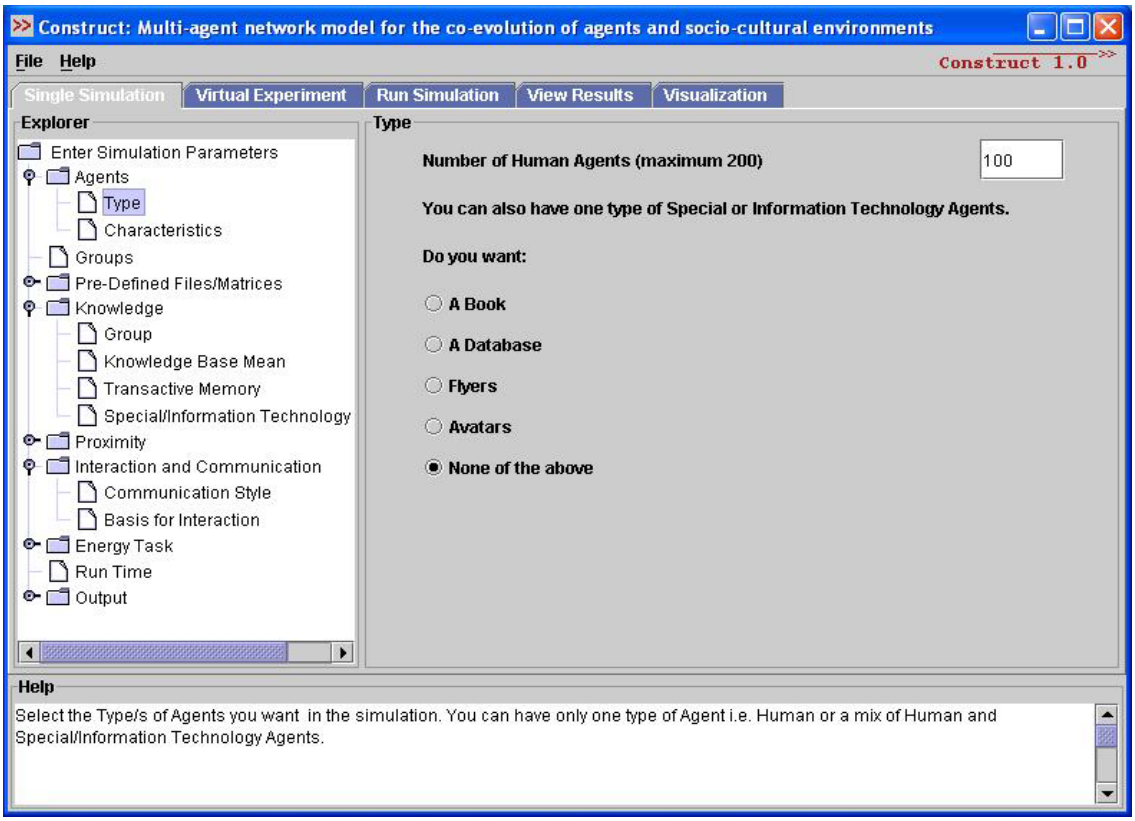
2. Model Input Variables and Output Files

On a system level, the construct model takes in groups/organizations and their characteristics as input. It then runs these groups/organizations through the Construct Interaction Cycle and generates system output. This process is also referred to as running a simulation.

2.1 Input Variables

Groups/Organizations and their characteristics are defined in terms of the following variables that constitute the Construct system input:

Figure 2: Construct Input Screen



- Agents
  - Type - Number of Human Agents
    This parameter is the total number of human agents in all the groups or organizations. The range for this parameter is 1 to 101, inclusive.

  - Type - Special or Information Technology Agent
    This parameter specifies (if there is any) the type of special or information technology agent – Book, Database, Flyer, Avatar.

  - Characteristics - Database
    This parameter specifies the characteristics of database technology – task based or referential. Task databases store facts related to the task. Referential databases store information about which agents know what facts.

  - Characteristics – Human Agents Forget
    This set of parameters specifies whether human agents forget and at what rate. The range for the forgetfulness parameter is from 0 to 1, inclusive. The forget rate determines how much of each agents knowledge base is forgotten each time period.

- Groups
  - Number of Groups
    This is the total number of groups. The range for this parameter is 1 to 3, inclusive.

- Number of Agents in Groups
  This set of parameters breaks down the number of agents into the various groups. The sum of these parameters must equal the total number of human agents plus any special or information technology agents.

- Predefined Matrices
  Construct allows the user to enter data (matrices) directly from pre-defined files. These files are either flat, unformatted files (*.txt) or UCINET files (*.dl). These allow the user fine control over knowledge and decisions and make it possible to input detailed real data. There are five types of matrices that can be input:

  - Requirements Network: This File is tasks by knowledge (facts) in size. This is a binary file such that Rik = 1 means for task i knowledge (fact) k bears positively on the decision. Rik = 0 means for task i knowledge k has no bearing on the decision. This file is also known as Decision File. Presently the number of tasks has to be 25. Thus there are task (25) rows and knowledge columns. Numbers are separated by a single space. Each row is ended by a carriage return.

  - Votes Network: This file is agents by knowledge in size. This is a numbers file filled with Vik: Vik = 1 means agent i feels that k bears positively on the decision. Vik = 0 means agent i feels that k has no bearing on the decision. Vik = -1 means agent i feels that k bears negatively on the decision. There are agent rows and knowledge columns. Numbers are separated by a single space. Each row is ended by a carriage return.

  - Knowledge Network: This file is agents by knowledge in size. This is a binary file such that: Kik=1 means agent i knows knowledge k. There are agent rows and knowledge columns. Numbers are separated by a single space. Each row is ended by a carriage return.

  - Transactive Knowledge Network: This file is agents by agents by knowledge in size. This is a binary file such that: Kijk = 1 means agent i thinks agent j knows k. There are agent matrices such that in each there are agent rows and knowledge columns. Each matrix is separated by a blank line. Numbers are separated by a single space. Each row is ended by a carriage return.

  - Access: This file is agents by agents in size. This is a numeric file such that: Rik = A means agent i has access to agent j, A percent of the time. A is greater than 0 less than 1. A = 1 means interaction is only controlled by knowledge. A = 0 means the two agent i cannot contact agent j directly. Rows indicate senders and columns indicate receivers. Numbers are separated by a single space. Each row is ended by a carriage return.

  To select a predefined file for input the user needs to check the box next to the predefined file type. Then select the format of the input file, either .txt or .dl, from the format drop-down menu. Finally, use the select file navigation button to search the local machines hard drive for the file and select it for input.

- Knowledge
  - Make Knowledge Fully Connected
    The parameter insures that every fact is known by at least one agent at the beginning of the simulation. Therefore all facts will be knowable initially. This parameter only works with random generation of the knowledge base and does not apply when a predefined knowledge network file is used for input.

  - Group - Number of Facts
    This parameter specifies the total number of facts (knowledge) across all the groups. The range for this parameter is 1 to 500, inclusive.

  - Group - Number of Facts special to Groups
    This set of parameters identifies facts that are special to the various groups. Special facts are assigned to the groups in sequential blocks. Group 1 will receive the first sequential block beginning with fact 1. Group 2's sequential block will begin where Group 1's block ended and so forth. Therefore special facts are not overlapping between groups. The sum of these parameters must not exceed total number of facts. See the Knowledge Base Mean for additional information on the implementation of special facts within the model. This parameter only works with random generation of the knowledge base and does not apply when a predefined knowledge network file is used for input.

  - Knowledge Base Mean
    This set of parameters determines on average what percentage of the group's knowledge (facts) is known by each agent at the beginning of the simulation. Facts are distributed randomly. The range for these parameters is 0 to 1, inclusive. These parameters only work with random generation of the knowledge base and do not apply when a predefined knowledge network file is used for input.

    - Percent of Facts special to Group i known by an agent in Group i
      These parameters distribute facts within each group. Each agent receives on average the percent of facts specified. Facts are distributed only within each group's respective block and the percentage applies to the block level. Use of the fully connected parameter may distribute some facts beyond the specified percentage that are outside of the respective blocks.

    - Percent of Facts special to Group j known by an agent in Group i
      These parameters distribute facts between each group. Each agent in group i receives on average the percent of facts specified. Facts are distributed from group j's block to agents in group i.

  - Transactive Memory
    This parameter distributes transactive memory to each agent. Transactive memory is the representation of agents i's understanding of what facts each of the other agents know. Each agent receives on average the percent of transactive memory specified at the beginning of the simulation. The percentage is applied at the aggregate group

level. The range for this parameter should be between 0.1 to 1.0, inclusive. This parameter only works with random generation of the transactive knowledge base and do not apply when a predefined transactive knowledge network file is used for input.

- Special/Information Technology
  These parameters distribute facts to each special/information technology agent at the beginning of the simulation. The book, flyer and avatar agents have a fixed knowledge base throughout the simulation so their percentage of facts known will not change. The database technologies will start with the initial percentage of facts known and their knowledge base will change as interaction with other agents occurs.

- Proximity
  This set of parameters determines if proximity influences the interactions between agents. To use proximity, check the Proximity Used box. Otherwise this box unchecked will ignore the effects of proximity. If proximity is used, the Proximity of Knowledge Base box allows for proximity to be calculated on the knowledge base rather than the concept of group closeness. All proximity measures are applied at the group level. Proximity is an additive in the relative measure calculation.

  - Proximity Mean
    These parameters determine the proximity effects within and between groups. The Mean for Proximity for Group i parameter determines how proximate each member of group i is to each other – intra-group. The Mean for Proximity for Group i to j determines how proximate each member of group i is to each member of group j – inter-group. The range for each proximity mean is 0 to 1, inclusive.

  - Proximity Variance
    These parameters determine the variance that is applied to each of the respective proximity means above. The range for each proximity variance is 0 to 1, inclusive.

  - Proximity Knowledge Base
    These parameters determine the knowledge base proximity effects. For each directional agent pair, the proximity value will be the multiplicative of the proximity knowledge base with the corresponding proximity mean. The proximity value of agents to themselves is the proximity mean. The range for each proximity knowledge base is 0 to 1, inclusive.

- Interaction and Communication
  These parameters determine human agent interaction and communication behaviors.

  - Communication Style
    This parameter determines whether human agents communicate only 1:1 each time period or have the ability to 1:N in a time period. Note: Agents can still communicate 1:1 with the 1:N style.

- Referral
  This parameter determines the use of referrals by the agents. Agents can receive referrals either from the transactive knowledge of other agents or from the referential database. When a referral is communicated, agent Y is instructed to go to agent Z for knowledge. Agent Y then attempts to follow up with agent Z in the next time period. Referrals ignore the effect of the relative measures, proximity and access except when access is set to 0. If access to the receiver agent is set to 0 then the sender agent is not allowed to contact the receiver agent. To use referrals check the Enable Referral box, otherwise referrals are inactive.

- Interaction Style
  This parameter determines the interaction style of the agents. For a description of the relative similarity and relative expertise styles see section 1.3, Interaction mechanism. If relative similarity or relative expertise is selected then these interaction styles are used exclusively throughout the simulation. To use both interaction styles within the same simulation select the Interaction is Both box and enter the percent of time that relative similarity is used. The percent usage for relative expertise in the use both style will be 1 – relative similarity percentage.

- Downsizing
  This set of parameters determine if, how and when downsizing occurs for a group.

  - Downsize
    To downsize a group select one of the Downsize Node options, otherwise select Do Not Downsize to turn downsizing off. There are five options for selecting a node for downsizing: highest cognitive load, highest centrality, highest betweenness centrality, highest closeness centrality and random selection. Every time a downsizing occurs, the downsized node is selected from the remaining set nodes according to the downsizing option chosen. The downsized node is removed from the remaining set of nodes without replacement. In addition, all facts known by that node are also removed.

  - Number of Time Periods between downsizing
    This parameter determines how often downsizing occurs. The downsizings occur at equally spaced intervals.

  - Number of Time Periods before downsizing begins
    This parameter determines at what time period downsizing starts.

  - Probability that an individual is missed in downsizing calculation
    This parameter determines the probability that a node is not downsized when a downsizing was supposed to have occurred.

  - Probability that a fact is missed in downsizing calculation
    This parameter determines the probability that a fact possessed by a downsized node remains in the groups memory after downsizing has occurred.

- Energy Task
  The energy task is used to replicate the amount of effort or time needed to complete tasks with varying complexities. Multitasking can be represented and the tradeoff between effort and accuracy can be analyzed. The energy task is separate from the binary choice task. In the energy task, agents are assigned tasks that are scheduled to appear at specified time periods throughout the simulation. Each task requires both knowledge specificity and amount for completion. Agents apply their knowledge to the task until both specificity and amount are satisfied which would represent full task completion and accuracy. Anything less than full satisfaction results in uncertainty and tasks that are not fully completed and less accurate. It is assumed that when agents are multitasking their effort and application of knowledge each time period is divided by the total number of tasks they currently are assigned. The effort and application of knowledge is thus evenly distributed to each task every time period.

    - Enable Energy Tasks
      To use the energy task check the Enable Energy Tasks box and input the total number of tasks that will be used in the simulation. This number should match the number of tasks that are used in constructing the task matrices described below.

  In addition to the knowledge base (People by Knowledge matrix) the energy tasks uses the following matrices:

    - Task by Agent
      This binary matrix represents the task assignment by agent. There can be multiple agents assigned to a task and multiple tasks assigned to an agent.

    - Task by Energy
      This non-binary matrix represents the knowledge that is required for the completion of a task. The non-binary string for each task shows the relative importance of each piece of knowledge to the task. A higher number indicates that this knowledge area requires more input into the task.

    - Task by Time Period
      This binary matrix represents the time period that each task gets introduced into the simulation. It is possible to have recurring tasks as a task can be introduced in more than one time period.

  To select a predefined matrix for input the user needs to check the box next to the predefined file type. The format of the input file should be .txt. Next, use the select file navigation button to search the local machines hard drive for the file and select it for input.

- Run Time
  This set of parameters determines the number of monte carlo runs, the number of time steps per run and the termination criteria for the simulation.

- Number of Runs:
  This is the number of runs over which results will be averaged. This parameter can take any positive integer value. As the value is increased the running time increases significantly (see performance characteristics section).

- Number of Time Steps
  This is the number of time steps per run. The range for this parameter is 1 to 1000, inclusive.

- Terminate Simulation
  This parameter determines when the simulation will terminate. The terminate options are: when all individuals know all facts (not recommended, can make run time exponential), when all agents in group 1 know all facts, when all agents except those in group 1 know all facts, when all agents except agent 1 in group 1 knows all facts and when the maximum number of time steps per run has been reached. The option to terminate when the maximum number time steps have been reached is the recommended option.

2.2 Output Files

To obtain a particular output check the format box under the respective output heading. Format options are file, screen or both. To obtain both output formats check both format boxes. If matrices are being output as files then select a file format, either .txt or .dl, under the Output tag in the interface. The following are summaries of the output options available in Construct. For details on the contents of each file refer to Appendix A or click the hyperlink to go directly to that files detail.

- Interactions (AgentInteractions.txt)
  This is a snapshot of who interacts with whom during every time period in the simulation. It is output as agent by agent matrices with a 1 indicating an interaction and a 0 indicating no interaction. The matrix rows indicate the sender and the matrix columns indicate the receiver.

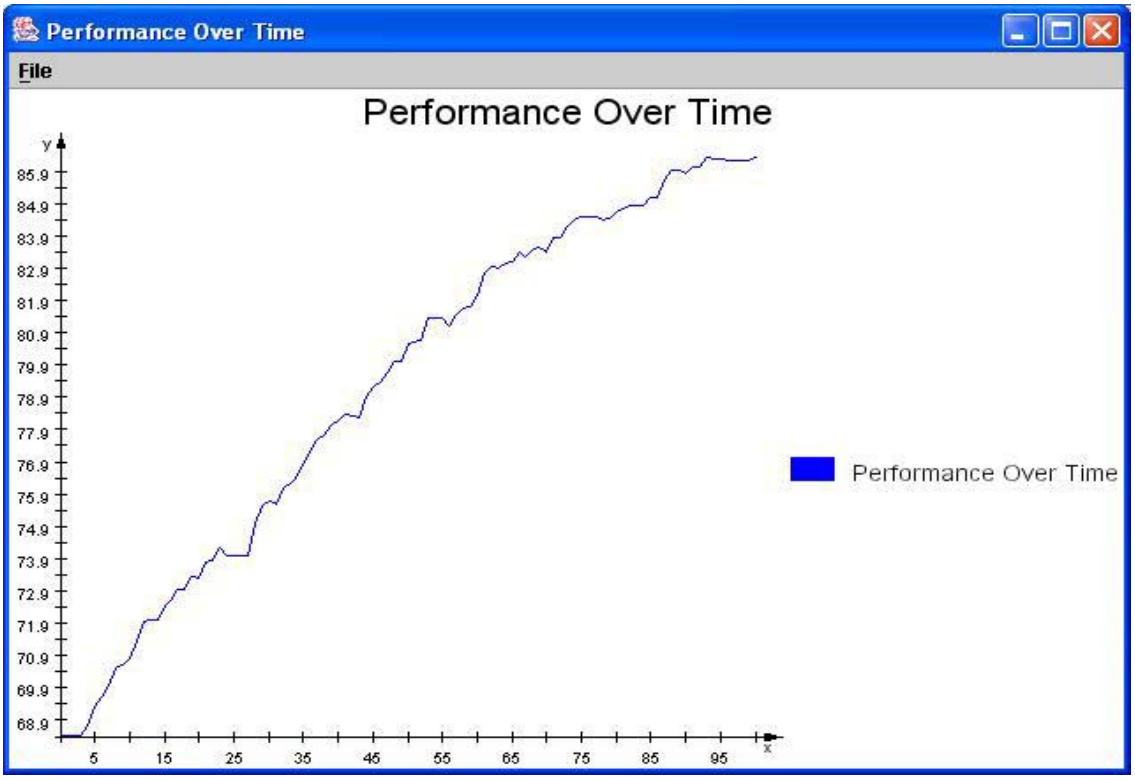
  This file is always output and is independent of any output flag setting.

- Output (output.txt)
  This file contains statistics on the number of agents and time periods, downsizing by time period, number of facts shared between agents, number of agents having consensus and number of agents having false consensus.

  This file is always output and is independent of any output flag setting.

- Performance Over Time (PerformanceTime.csv)
  This is the overall performance accuracy of all agents for every time period averaged over the number of runs.

Figure 3: Construct – Performance Output Plot



- Diffusion Over Time (DiffusionTime.csv)
  This is the diffusion of the first fact among all agents for every time period averaged over the number of runs.

- Average Energy per Time Period (AvgEnergyTime.csv)
  This file contains a measure of the average amount of resources (energy) available to perform all tasks for each time period.

- Average Sufficient Effort per Time Period (AvgSuffEffortTime.csv)
  This file contains a measure of the average amount of effort (energy) applied to all tasks that is sufficient for the tasks to be performed accurately. This measure is given for each time period.

- Friends and Enemies – Interaction Thirds (FriendsEnemiesThirds.csv)
  This file contains measures of Friends and Enemies based on interaction thirds and Levine's study.

- Friends and Enemies – Levine's Study (FriendsEnemiesLevine.csv)
  This file contains measures of Friends and Enemies based on Levine's study.

- Transactive Memory (TransactiveMem.csv)
  This file contains transactive memory measures along with measures of referrals, diffusion, performance accuracy, energy, sufficient effort and number of agents. Each of the measures is given by time period. Also, comparisons are made to an illustrative task of size 10 for the performance accuracy, energy and sufficient expertise measures. The illustrative task assumes a knowledge base of size 50.

- Baysian T-test (BaysianT.csv)
  This file contains values pertaining to Bayesian T-Test. It is printed once every run. This output is currently disabled.

- Simulation Setup (GroupStructure.csv)
  This file contains the initial group structure for simulation.

- Probability of Interaction (ProbInteraction.txt)
  This is a snapshot of the probability of interaction between every pair of agents by time period. It is output as agent by agent matrices with the rows indicating the sender and the columns indicating the receiver.

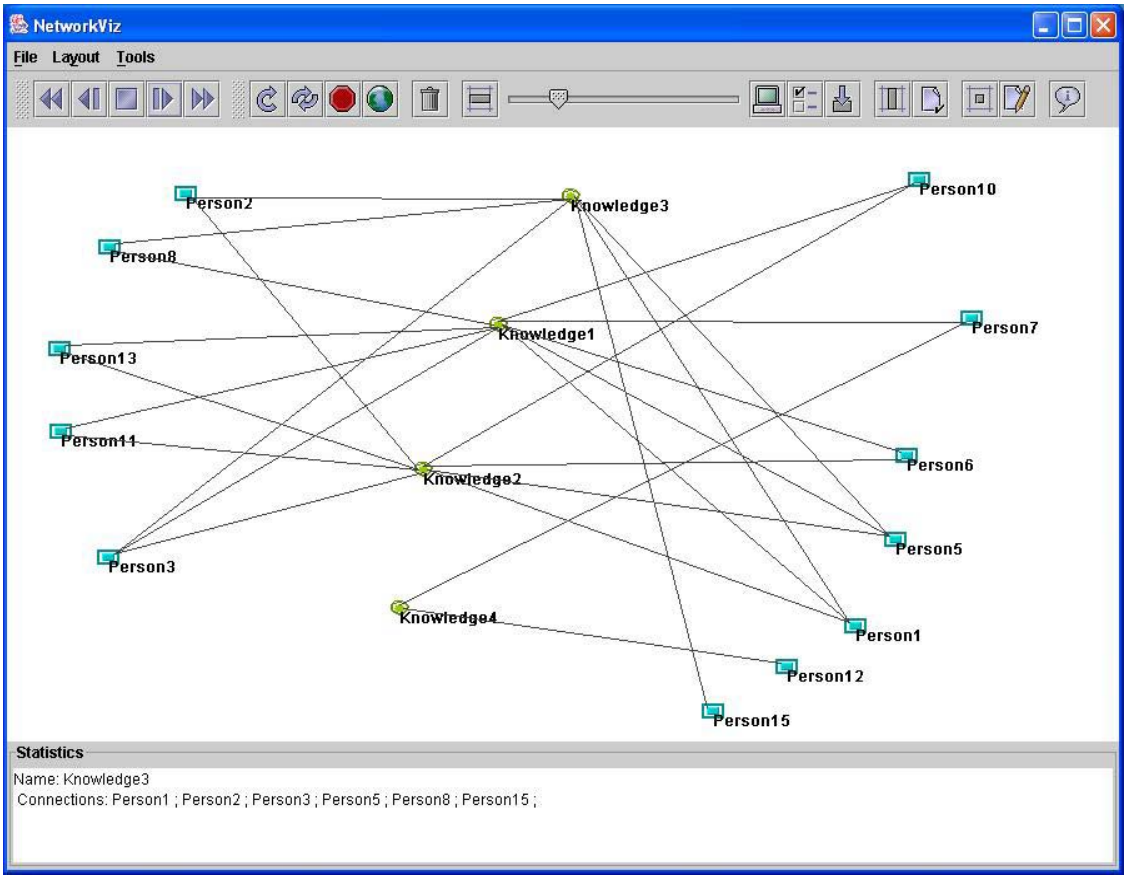
- Proximity (Proximity.txt)
  This is an agent by agent matrix of the proximity value for each dyad. This matrix is symmetric and static throughout the simulation.

- Friendship Network (FriendshipNets.txt)
  This file contains valued agent by agent matrices by time period. The values in the matrices represent relative measures of ties between agents based on friendship.

- Advice Network (AdviceNets.txt)
  This file contains valued agent by agent matrices by time period. The values in the matrices represent relative measures of ties between agents based on advice seeking.

- Knowledge Base (KnowledgeBase.txt)
  This is a snapshot of the knowledge base of the organization during every time period in the simulation. It is output as an agent by knowledge (facts) matrix with 1 indicating that the agent knows that piece of knowledge (fact) 0 indicating that the agent does not know the piece of knowledge (fact).

Figure 4: Construct – Knowledge Base Visualization



- Transactive Knowledge Base (TrKnowledgeBase.txt)
  This is a snapshot of the transactive knowledge base of the organization during every time period in the simulation. There are agent matrices such that in each there are agent rows and knowledge columns. A 1 indicates that an agent knows that the corresponding agent knows that piece of knowledge (fact) and a 0 indicates that an agent does not know if the corresponding agent knows that piece of knowledge (fact).

- Database Parameters (Database.csv)
  This file contains data related to database use during the simulation. Data include agent contributions and retrievals, novel contributions and retrievals, referential data contributions and retrievals, task data contributions and retrievals, average contributor expertise and average retriever expertise.

- Summary (Summary.csv)
  This file contains measures of average initial and final performance accuracy, diffusion, energy and sufficient effort. It also has differential measures for comparing the effects of downsizing on performance accuracy, diffusion, energy and sufficient effort.

- Time Series (TimeSeries.csv)
  This is time series data from the first iteration run. Measures of interaction, triads, diffusion, performance accuracy, consensus and cultural homogeneity are printed given for every time period.

- Performance/Diffusion Comparison (PerfDiffCompare.csv)
  This is a comparison of performance and diffusion for every time period. These measures are the same measures used the Performance Over Time and Diffusion Over Time output files. The values are averaged over the number of runs.

- Average Triads (AvgTriads.csv)
  This file contains a measure of the average general triads among all agents by time period.

3. System Requirements

The typical configuration of a machine running Construct would be:

OS Name	Microsoft Windows XP Professional
Version	5.1.2600 Service Pack 1 Build 2600
System Type	X86-based PC
Processor	x86 Family 6 Model 9 Stepping 5 GenuineIntel ~1594 Mhz
Total Physical Memory	512.00 MB
Available Physical Memory	208.91 MB
Total Virtual Memory	2.97 GB
Available Virtual Memory	2.41 GB
Page File Space	2.47 GB

4. Performance Characteristics

Based on the configuration specified above below are the performance characteristics for certain simulation samples (Table 2):

Table 2: Construct performance characteristics

Agents	Facts	Time Periods	Runs	Output Type	Running Time	Output Size
19	57	100	5	All	14 seconds	23.5 Mb
19	57	100	5	Only Performance, Diffusion and Interactions	6 seconds	472Kb
19	57	100	50	All	1 min 48 sec	228 Mb
19	57	100	50	Only Performance, Diffusion and Interactions	56 sec	4 Mb

Note:

1) CPU usage is almost 100 % when a Construct simulation is running

2) The Knowledge Base output file and especially the Transactive Knowledge Base output file requires considerable storage space for simulations with a large number of runs or time periods e.g. for a 19 agent, 57 fact, 100 time periods, 50 runs simulation the size of the knowledge base file is 10 Mb and the size of the corresponding transactive knowledge base file is 200 Mb.

5. Illustrative Examples

The following sections show different ways to model a group of five agents with ten total facts available. Each agent knows three facts.

5.1 Model random generation

The model will randomly assign facts to each agent.

Under the File menu

1) Set Working Directory
    a) Set the working directory to where you want to store the output files

Under the Single Simulation tab

2) Agents – Type
    a) Number of agents = 5
3) Groups
    a) Number of agents in group 1 = 5
4) Knowledge – Group
    a) Total number of facts = 10
    b) Number of facts special to group1 = 10
5) Knowledge – Knowledge base mean
    a) Percent of facts known to an agent in group 1 = 0.30
6) Knowledge – Transactive memory
    a) Percent of transactive memory that agents have = 1.0
7) Interaction and Communication
    a) Make sure communication style = 1:1
    b) Make sure basis for interaction = relative similarity
8) Run Time
    a) Number of runs to average = 5
    b) Number of time steps per run = 50
9) Output – Screen/File – Time
    a) Performance over time
        i) Leave Write to output file checked
    b) Diffusion over time
        i) Leave Write to output file checked
    c) Uncheck all other options under the Output – Screen/File tab

Under the Run Simulation tab

10) Select Run Options
    a) Run single simulation – click on RUN

Under the View Results tab

11) Performance Over Time
    a) On the left-side explorer menu highlight Performance Over Time
    b) Click on Generate
12) Diffusion Over Time
    a) On the left-side explorer menu highlight Diffusion Over Time
    b) Click on Generate

5.2 Create a knowledge base for input

Suppose we want to represent a known distribution such as the following whereas a 1 represents an agent that knows that bit of knowledge and a 0 represents that an agent does not know that bit of knowledge

	Knowledge
A	1 1 1 0 0 0 0 0 0 0
g	0 0 1 1 1 0 0 0 0 0
e	0 0 0 0 1 1 1 0 0 0
n	0 0 0 0 0 0 1 1 1 0
t	0 0 0 0 0 0 0 1 1 1

1) Copy the above knowledge base into a text file (excluding the column and row labels) and name it Testkb.txt

Under the File menu

2) Set Working Directory
    a) Set the working directory to where you want to store the output files
    b) Copy Testkb.txt into this directory

Under the Single Simulation tab

3) Agents – Type
    a) Number of agents = 5
4) Groups
    a) Number of agents in group 1 = 5
5) Pre-Defined Files/Matrices
    a) Select Knowledge Network
        i) Check the box for Knowledge Network File in the frame on the right
    b) Make sure txt is selected as the format
    c) Select Choose File

i) Select Testkb.txt
i) Click Open

6) Knowledge – Group
    a) Total number of facts = 10
    a) Number of facts special to group1 = 10

Under the Run Simulation tab

7) Select Run Options
    a) Run single simulation – click on RUN

Under the View Results tab

8) Performance Over Time
    a) On the left-side explorer menu highlight Performance Over Time
    b) Click on Generate

9) Diffusion Over Time
    a) On the left-side explorer menu highlight Diffusion Over Time
    b) Click on Generate

References

Carley, K. M. An approach for related social structure to cognitive structure. Journal of Mathematical Sociology, v.12, n.2, p.137-189. 1986a.

_____. Knowledge acquisition as a social phenomenon. Instructional Science, v.14, p.381-438. 1986b.

_____. Group stability: A socio-cognitive approach. In: Lawler E., Markovsky, B., *et al* (Ed.). Advances in group processes: Theory & research. Greenwhich, CN: JAI Press, v.VII, 1990. Group stability: A socio-cognitive approach, p.1-44

_____. Growing up: The development and acquisition of social knowledge. In: Howard, J. and Callero. P. (Ed.). The Self-Society Dynamic: Cognition, Emotion, and Action. Cambridge, England: Cambridge University Press, 1991a. Growing up: The development and acquisition of social knowledge, p.72-105

_____. A theory of group stability. American Sociological Review, v.56, n.3, p.331-354. 1991b.

_____. Communication technologies & their effect on cultural homogeneity, consensus, & the diffusion of new ideas. Sociological Perspectives, v.38, n.4, p.547-571. 1995.

_____. Communicating new ideas: The potential impact of information and telecommunication technology. Technology in Society, v.18, n.2, p.219-230. 1996.

_____. On the evolution of social and organizational networks. In: S. B. Andrews and D. Knoke (Ed.). Special issue of Research in the Sociology of Organizations on Networks In and Around Organizations. Greenwhich, CN: JAI Press, v.16, 1999. On the evolution of social and organizational networks, p.3-30

_____. Smart agents and organizations of the future. In: L. Lievrouw and S. Livingstone (Ed.). The Handbook of New Media. Thousand Oaks, CA: Sage, 2002. Smart agents and organizations of the future, p.206-220

Carley, K. M. and V. Hill. Structural change and learning within organizations. In: A. Lomi and. E. R. Larsen (Ed.). Dynamics of Organizations: Computational Modeling and Organization Theories. Menlo Park, CA: MIT Press/AAAI, 2001. Structural change and learning within organizations

Carley, K. M. and D. Krackhardt. Cognitive inconsistencies and non-symmetric friendship. Social Networks, v.18, p.1-27. 1996.

Carley, K. M. and C. Schreiber. Information technology and knowledge distribution in C3I teams. 2002 Command and Control Research and Technology Symposium. Monterey, CA, 2002. p.

Constant, D., S. L., *et al.* The kindness of strangers: The usefulness of electronic weak ties for technical advice. Organization Science, v.7, n.2, p.119-135. 1996.

Festinger, L. Informal social communication. Psychology Review, v.57, p.271-282. 1950.

_____. A theory of social comparison processes. Human Relations, v.7, p.117-140. 1954.

Festinger, L., S. Schachter, *et al.* Social pressures in informal groups: A study of human factors in housing. Palo Alto, CA: Stanford University Press. 1950

Giddens, A. The constitution of society: Outline of the theory of structuration. Berkeley, CA: University of California Press. 1984

Grannovetter, M. Getting a job: A study of contacts and careers. Cambridge: Harvard University Press. 1974

Hollenbeck, J., D. Ilgen, *et al.* Team performance on monitoring tasks: An examination of decision errors in contexts requiring sustained attention. Journal of Applied Psychology, v.80, p.685-696. 1995.

Kaufer, D. S. and K. M. Carley. Communication at a Distance: The Effect of Print on Socio-Cultural Organization and Change. Hillsdale, NJ: Lawrence Erlbaum Associates. 1993

Lazarsfeld, P. and R. Merton. Friendship as a social process: A substantive and methodological analysis. In: M. Berger, T. Abel, *et al* (Ed.). Freedom and control in modern society. New York: Octagon, 1954. Friendship as a social process: A substantive and methodological analysis

Manis , J. G. and B. N. Meltzer. Symbolic interaction: A reader in social psychology. Boston: Allyn & Bacon. 1978

Mcpherson, J. M. and L. Smith-Lovin. Homophily in voluntary organizations: Status distance and the composition of face-to-face groups. American Sociological Review, v.52, p.370-379. 1987.

Merton, R. Continuities in the theory of reference groups and social structure. In: R. Merton (Ed.). Social theory and social structure. New York: Free Press, 1968. Continuities in the theory of reference groups and social structure

Moreland, R., L. Argote, *et al.* Socially shared cognition at work: Transactive memory and group performance. In: J. Nye and A. Bower (Ed.). What's social about social cognition? Thousand Oaks, CA: Sage, 1996. Socially shared cognition at work: Transactive memory and group performance, p.57-84

Salancik, G. R. and J. Pfeffer. A social information processing approach to job attitudes and task design. Administrative Science Quarterly, v.23, p.224-253. 1978.

Schreiber, C. and K. M. Carley. The impact of databases on knowledge transfer: Simulation providing theory. 2003 NAACSOS conference proceedings. Pittsburgh, PA, 2003.

_____. Going beyond the data: Empirical validation leading to grounded theory. Computational and Mathematical Organization Theory. in press.

Wegner, D. M. Transactive memory: A contemporary analysis of the group mind. In: B. Mullen and G. R. Goethals (Ed.). Theories of group behavior. New York: Springer-Verlag, 1986. Transactive memory: A contemporary analysis of the group mind, p.185-205

Appendix A
Output file detail

(To return to the file summary of the main document - click the filename hyperlink)

- Interactions (AgentInteractions.txt)
  This is a snapshot of who interacts with whom during every time period in the simulation. It is output as agent by agent matrices with a 1 indicating an interaction and a 0 indicating no interaction. The matrix rows indicate the sender and the matrix columns indicate the receiver.

  This file is always output and is independent of any output flag setting.

- Output (output.txt)
  This file contains statistics on the number of agents and time periods, downsizing by time period, number of facts shared between agents, number of agents having consensus and number of agents having false consensus.

  Number of agents (not listed by time period)
  All time periods executed including the agent downsizing information in the following format:
  "setting skb for individual 0 to zero at time 40"
  This means that agent 0 has been eliminated (organizational downsizing) from the simulation at time period 40. Note that agents are numbered from 0 to n-1 where n is the total number of agents in the simulation both human and artificial.
  Average initial knowledge base and average initial proximity parameters are also provided.
  Number of facts shared between agents
  Column 1: Time period
  Column 2: Average number of all facts shared between all agents
  Column 3: Standard deviation of all fact shared between all agents
  Column 4: Average number of knowledge links
  Column 5: Standard deviation of knowledge links
  Column 6: Average number of all facts shared between agents in group 1
  Column 7: Standard deviation of all fact shared between agents in group 1
  Column 8: Average number of all facts shared between agents in group 2
  Column 9: Standard deviation of all fact shared between agents in group 2
  Column 10: Average number of all facts shared between agents in group 3
  Column 11: Standard deviation of all fact shared between agents in group 3
  Number of agents sharing all facts with agent 1 in group 1
  Column 1: Time period
  Column 2: Average number of all agents sharing all facts with agent 1 in group 1
  Column 3: Standard deviation of all agents sharing all facts with agent 1 in group 1
  Column 4: Average number of group 1 agents sharing all facts with agent 1 in group 1
  Column 5: Standard deviation of group 1 agents sharing all facts with agent 1 in group 1
  Column 6: Average number of group 2 agents sharing all facts with agent 1 in group 1

Column 7: Standard deviation of group 2 agents sharing all facts with agent 1 in group 1
Column 8: Average number of group 3 agents sharing all facts with agent 1 in group 1
Column 9: Standard deviation of group 3 agents sharing all facts with agent 1 in group 1

Number of agents sharing all at lease one fact with agent 1 in group 1

Column 1: Time period
Column 2: Average number of all agents sharing one fact with agent 1 in group 1
Column 3: Standard deviation of all agents sharing one fact with agent 1 in group 1
Column 4: Average number of group 1 agents sharing one fact with agent 1 in group 1
Column 5: Standard deviation of group 1 agents sharing one fact with agent 1 in group 1
Column 6: Average number of group 2 agents sharing one fact with agent 1 in group 1
Column 7: Standard deviation of group 2 agents sharing one fact with agent 1 in group 1
Column 8: Average number of group 3 agents sharing one fact with agent 1 in group 1
Column 9: Standard deviation of group 3 agents sharing one fact with agent 1 in group 1

Number of agents having consensus

Column 1: Time period
Column 2: Average number of all agents having consensus
Column 3: Standard deviation of all agents having consensus
Column 4: Average number of group 1 agents having consensus
Column 5: Standard deviation of group 1 agents having consensus
Column 6: Average number of group 2 agents having consensus
Column 7: Standard deviation of group 2 agents having consensus
Column 8: Average number of group 3 agents having consensus
Column 9: Standard deviation of group 3 agents having consensus

Number of agents having false consensus

Column 1: Time period
Column 2: Average number of all agents having false consensus
Column 3: Standard deviation of all agents having false consensus
Column 4: Average number of group 1 agents having false consensus
Column 5: Standard deviation of group 1 agents having false consensus
Column 6: Average number of group 2 agents having false consensus
Column 7: Standard deviation of group 2 agents having false consensus
Column 8: Average number of group 3 agents having false consensus
Column 9: Standard deviation of group 3 agents having false consensus

This file is always output and is independent of any output flag setting.

- Performance Over Time (PerformanceTime.csv)
  This is the overall performance accuracy of all agents for every time period averaged over the number of runs.

  Row 1: Average performance accuracy. The first column is the average for time period 1, the second column is the average for time period 2 …

- Diffusion Over Time (DiffusionTime.csv)
  This is the diffusion of the first fact among all agents for every time period averaged over the number of runs.

Row 1: Average diffusion of the first fact. The first column is the average for time period 1, the second column is the average for time period 2 …

- Average Energy per Time Period (AvgEnergyTime.csv)
  This file contains a measure of the average amount of resources (energy) available to perform all tasks for each time period.

  Column 1: Time period
  Column 2: Average energy available

- Average Sufficient Effort per Time Period (AvgSuffEffortTime.csv)
  This file contains a measure of the average amount of effort (energy) applied to all tasks that is sufficient for the tasks to be performed accurately. This measure is given for each time period.

  Column 1: Time period
  Column 2: Average amount of sufficient energy applied

- Friends and Enemies – Interaction Thirds (FriendsEnemiesThirds.csv)
  This file contains measures of Friends and Enemies based on interaction thirds and Levine's study.

- Friends and Enemies – Levine's Study (FriendsEnemiesLevine.csv)
  This file contains measures of Friends and Enemies based on Levine's study.

- Transactive Memory (TransactiveMem.csv)
  This file contains transactive memory measures along with measures of referrals, diffusion, performance accuracy, energy, sufficient effort and number of agents. Each of the measures is given by time period. Also, comparisons are made to an illustrative task of size 10 for the performance accuracy, energy and sufficient expertise measures. The illustrative task assumes a knowledge base of size 50.

  Column 1: Average accuracy of all agents' transactive memory
  Column 2: Average amount of all agents' transactive memory
  Column 3: Average number of additions to transactive memory
  Column 4: Percentage of agents who get referrals
  Column 5: Percentage of agents who follow referrals
  Column 6: Percentage of agents who receive referrals from a database
  Column 7: Percentage of agents who follow referrals from a database
  Column 8: Average diffusion of the first fact (simple message) among all agents
  Column 9: Average diffusion of facts 1 through 5 (complex message) among all agents
  Column 10: Average performance accuracy for all agents
  Column 11: Percent improvement of average performance accuracy as compared to guessing

Column 12: Average performance accuracy for an illustrative task size of 10
Column 13: Percent improvement of average performance accuracy as compared to guessing for an illustrative task size of 10
Column 14: Average energy
Column 15: Average energy for an illustrative task size of 10
Column 16: Sufficient effort
Column 17: Sufficient effort for an illustrative task size of 10
Column 18: Average number of agents

- Baysian T-test (BaysianT.csv)
  This file contains values pertaining to Bayesian T-Test. It is printed once every run. This output is currently disabled.

- Simulation Setup (GroupStructure.csv)
  This file contains the initial group structure for simulation.

  Row 1: Number of groups
  Row 2: Number of agents in group1
  Row 3: Number of facts for group 1
  Row 4: Number of agents in group 2
  Row 5: Number of facts for group 2
  Row 6: Number of agents in group 3
  Row 7: Number of facts for group 3

- Probability of Interaction (ProbInteraction.txt)
  This is a snapshot of the probability of interaction between every pair of agents by time period. It is output as agent by agent matrices with the rows indicating the sender and the columns indicating the receiver.

- Proximity (Proximity.txt)
  This is an agent by agent matrix of the proximity value for each dyad. This matrix is symmetric and static throughout the simulation.

- Friendship Network (FriendshipNets.txt)
  This file contains valued agent by agent matrices by time period. The values in the matrices represent relative measures of ties between agents based on friendship.

- Advice Network (AdviceNets.txt)
  This file contains valued agent by agent matrices by time period. The values in the matrices represent relative measures of ties between agents based on advice seeking.

- Knowledge Base (KnowledgeBase.txt)
  This file contains a snapshot of the knowledge base of the organization during every time period in the simulation. It is output as an agent by knowledge (facts) matrix with 1 indicating that the agent knows that piece of knowledge (fact) 0 indicating that the agent does not know the piece of knowledge (fact).

- Transactive Knowledge Base (TrKnowledgeBase.txt)
  This file contains a snapshot of the transactive knowledge base of the organization during every time period in the simulation. There are agent matrices such that in each there are agent rows and knowledge columns. A 1 indicates that an agent knows that the corresponding agent knows that piece of knowledge (fact) and a 0 indicates that an agent does not know if the corresponding agent knows that piece of knowledge (fact).

- Database Parameters (Database.csv)
  This file contains data related to database use during the simulation. Data include agent contributions and retrievals, novel contributions and retrievals, referential data contributions and retrievals, task data contributions and retrievals, average contributor expertise and average retriever expertise.
  Column 1: Time period
  Column 2: Overall amount of novel facts contributed to the database
  Column 3: Overall total amount of facts contributed to the database
  Column 4: Overall number of agents who contribute to the database
  Column 5: Amount of novel referential data contributed to the database
  Column 6: Total amount of referential data contributed to the database
  Column 7: Number of agents who contribute referential data to the database
  Column 8: Amount of novel task data contributed to the database
  Column 9: Total amount of task data contributed to the database
  Column 10: Number of agents who contribute task data to the database
  Column 11: Overall number of agents who retrieve novel facts from the database
  Column 12: Overall number of agents who retrieve facts from the database
  Column 13: Overall number of agents who retrieve facts from the database
  Column 14: Number of agents who retrieve novel referential data from the database
  Column 15: Number of agents who retrieve referential data from the database
  Column 16: Number of agents who retrieve referential data from the database
  Column 17: Number of agents who retrieve novel task data from the database
  Column 18: Number of agents who retrieve task data from the database
  Column 19: Number of agents who retrieve task data from the database
  Column 20: Average expertise of contributor
  Column 21: Average expertise of retriever

- Summary (Summary.csv)
  This file contains measures of average initial and final performance accuracy, diffusion, energy and sufficient effort. It also has differential measures for comparing the effects of downsizing on performance accuracy, diffusion, energy and sufficient effort.

  Column1: Number of agents in the simulation
  Column2: Downsizing strategy
  Column3: Percent of agents known
  Column4: Percent of knowledge ties known
  Column5: Average initial performance accuracy (over the first 10 time periods)
  Column6: Average initial diffusion of the first fact (over the first 10 time periods)
  Column7: Average initial average energy (over the first 10 time periods)

Column8: Average initial average sufficient effort (over the first 10 time periods)
Column9: Average performance accuracy - response to first downsizing, time periods 51-61
Column10: Average diffusion of the first fact - response to first downsizing, time periods 51-61
Column11: Average energy - response to first downsizing, time periods 51-61
Column12: Average sufficient effort - response to first downsizing, time periods 51-61
Column13: Average final performance accuracy (over the last 10 time periods)
Column14: Average final diffusion of the first fact (over the last 10 time periods)
Column15: Average final energy (over the last 10 time periods)
Column16: Average final sufficient effort (over the last 10 time periods)
Column17: Percent difference in overall performance accuracy
Column18: Percent difference in overall diffusion of the first fact
Column19: Percent difference in overall average energy
Column20: Percent difference in overall average sufficient effort
Column21: Percent difference in performance accuracy due to downsizing
Column22: Percent difference in diffusion of the first fact due to downsizing
Column23: Percent difference in average energy due to downsizing
Column24: Percent difference in average sufficient effort due to downsizing

- Time Series (TimeSeries.csv)
  This is time series data from the first iteration run. Measures of interaction, triads, diffusion, performance accuracy, consensus and cultural homogeneity are printed given for every time period.

  INT: Interaction
  Column2: Average interaction among the agents
  Column3: Interaction standard deviation
  TRIADS:
  Column5: A measure of the general triads among all agents
  Column6: A measure of the strong triads among all agents
  Column7: A measure of the general triads among group 1 agents
  Column8: A measure of the general triads among group 2 agents
  Column9: A measure of the general triads between group 1 and group 2 agents
  DIFU: Diffusion
  Column11: Diffusion of the first fact
  Column12: Diffusion of facts 1 through 5
  Column13: Diffusion of the first fact within group 1
  Column14: Diffusion of the first fact within group 2
  A: Performance accuracy for all agents (overall group level)
  Column16: Total performance accuracy for a task using all facts
  Column17: Average performance accuracy for a task using all facts
  Column18: Total performance accuracy for a task using facts 1 through 5
  Column19: Average performance accuracy for a task using facts 1 through 5
  Column20: Total performance accuracy for a task using facts 1 through 10
  Column21: Average performance accuracy for a task using facts 1 through 10
  Column22: Total performance accuracy for a task using a sample of 5 facts from the total

number of facts
Column23: Average performance accuracy for a task using a sample of 5 facts from the total number of facts
Column24: Total performance accuracy for a task using a sample of 10 facts from the total number of facts
Column25: Average performance accuracy for a task using a sample of 10 facts from the total number of facts

A-G1: Performance accuracy for group 1
Column27: Group 1 Total performance accuracy for a task using all facts
Column28: Group 1 Average performance accuracy for a task using all facts
Column29: Group 1 Total performance accuracy for a task using facts 1 through 5
Column30: Group 1 Average performance accuracy for a task using facts 1 through 5
Column31: Group 1Total performance accuracy for a task using facts 1 through 10
Column32: Group 1 Average performance accuracy for a task using facts 1 through 10
Column33: Group 1 Total performance accuracy for a task using a sample of 5 facts from the total number of facts
Column34: Group 1 Average performance accuracy for a task using a sample of 5 facts from the total number of facts
Column35: Group 1 Total performance accuracy for a task using a sample of 10 facts from the total number of facts
Column36: Group 1 Average performance accuracy for a task using a sample of 10 facts from the total number of facts

A-G2: Performance accuracy for group 2
Column38: Group 2 Total performance accuracy for a task using all facts
Column39: Group 2 Average performance accuracy for a task using all facts
Column40: Group 2 Total performance accuracy for a task using facts 1 through 5
Column41: Group 2 Average performance accuracy for a task using facts 1 through 5
Column42: Group 2Total performance accuracy for a task using facts 1 through 10
Column43: Group 2 Average performance accuracy for a task using facts 1 through 10
Column44: Group 2 Total performance accuracy for a task using a sample of 5 facts from the total number of facts
Column45: Group 2 Average performance accuracy for a task using a sample of 5 facts from the total number of facts
Column46: Group 2 Total performance accuracy for a task using sample of 10 facts from the total number of facts
Column47: Group 2 Average performance accuracy for a task using a sample of 10 facts from the total number of facts

PA: Consensus
Column49: Consensus among all agents.
Column50: Consensus among group 1 agents.
Column51: Consensus among group 2 agents.

K: Cultural Homogeneity
Column53: Average cultural homogeneity for all agents.
Column54: Cultural homogeneity standard deviation.
Column55: Average cultural homogeneity for group 1.
Column56: Average cultural homogeneity for group 2.

- Performance/Diffusion Comparison (PerfDiffCompare.csv)
  This file contains measures of performance and diffusion for every time period. These measures are the same measures used the Performance Over Time and Diffusion Over Time output files. The values are averaged over the number of runs.

  Column 1: Average performance accuracy. The first row is the average for time period 1, the second row is the average for time period 2 …
  Column 2: Average diffusion of the first fact. The first row is the average for time period 1, the second row is the average for time period 2 …

- Average Triads (AvgTriads.csv)
  This file contains a measure of the average general triads among all agents by time period.

  Column 1: Time period
  Column 2: Average general triads