Enhancing the Credibility of Agent-Based Model for the Study of Workers' Group Behavior by Comparing Simulation Data with Survey Data

Seungjun Ahn[1] and SangHyun Lee[2]

[1] PhD candidate, Department of Civil & Environmental Engineering, University of Michigan, 1316 G. G. Brown, 2350 Hayward Street, Ann Arbor, MI 48109; email: esjayahn@umich.edu
[2] Assistant Professor, Department of Civil & Environmental Engineering, University of Michigan, 2340 G. G. Brown, 2350 Hayward Street, Ann Arbor, MI 48109; email: shdpm@umich.edu

ABSTRACT

Construction workers' behavior is an important factor of productivity and safety on a job site. Worker behavior is not only determined by individual's inherent characteristics, but also largely shaped by contexts which include the social interactions within a project. As a result, construction workers' behavior on a job site is under the influence of the social norms and work culture of their workgroup. However, our current understanding of the social aspect of worker behavior is limited. To expand our understanding of workers' social behavior and its impact on project performance, researchers have begun using an agent-based modeling approach because it allows researchers to observe the complex group-level behavior emerging from individuals' interactions. One of the important issues in using agent-based models for organizational research is enhancing the credibility of model predictions. With this background in mind, the objective of this paper is to propose a methodology that can help enhance the credibility of an agent-based model for the study of workers' social behavior using survey data. Specifically, it is suggested in this paper that the perceptual/attitudinal/behavioral data from surveys and the corresponding data from simulations be transformed into categorical data, and then compared to each other to show a quantitative agreement of the model behavior with empirical data. The proposed methodology is illustrated by an example of construction workers' absenteeism research that we have conducted. With the results of this research, it is argued that demonstrating the correspondence of simulation data to survey data using a data categorization method is an effective means to enhance the credibility of an agent-based model.

INTRODUCTION

Labor is a crucial resource in construction, and therefore construction workers' behavior significantly impacts overall project performance. A large number of previous studies show that workers' behavior influence project performance in various ways. Researchers working on labor productivity have found that "productivity of labor itself" is a major factor of a project's productivity (Maloney

1983). Researchers working on construction safety have also found that one of the major factors of construction accidents is workers' unsafe behavior (Heinrich et al. 1980; Hinze et al. 2005). Therefore, managing worker behavior on a job site should be an important element in construction management.

To expand our understanding of worker behavior, many researchers have studied the factors that affect worker behavior in construction. In these efforts, researchers have found that various personal and contextual factors affect workers' behavior that is associated with labor productivity and/or safety. Examples of the factors that have been revealed to affect labor productivity are personality, age, skill level, experience, contractual agreements, labor availability, union activities, and rework (Hendrickson 2000). Also, examples of the factors that have been revealed to affect workers' safety behavior are workers' safety awareness, work pressure, co-workers' attitudes (Choudhry and Fang 2008), and various working condition-related factors, such as required tools or equipment and layout conditions (Chi et al. 2012).

More recently, researchers have found that social norms and culture in workgroups also significantly affect construction workers' behavior concerning safety (Mitropoulos and Memarian 2012; Choudhry and Fang 2008; Mohamed 2002) and productivity (Ahn et al. 2013b; Maloney 1983). This is not surprising; as a social being like any of us, construction workers observe others, learn how to work by watching others, mimic others' behavior, and internalize the social norms of their societies (Bandura 1991). The social sciences including organizational behavior and psychology provide abundant theories and evidence that employees' behavior is shaped largely by the social mechanisms—such as social identity and social influences from co-workers (Friedkin 2004; Ashforth and Mael 1989)—and construction workers should not be an exception to these principles.

As more attention is being paid to the role of informal rules and social interactions in organizations, agent-based modeling (ABM) has emerged as a useful tool for the study of complex organizational phenomena because it allows researchers to observe the behavior of organization emerging from individuals' interactions. However, using ABM in organizational research is not an easy undertaking. One of the important issues in using agent-based models for organizational research is enhancing the credibility of model predictions.

In order to address this issue, we propose a methodology that can help enhance the credibility of agent-based models for the study of workers' social behavior. Specifically, it is suggested in this paper that the perceptual/attitudinal/behavioral data from surveys and the corresponding data from simulations be transformed into categorical data, and then compared to each other to show a quantitative agreement of the model behavior with empirical data. The proposed methodology is then illustrated by an example of construction workers' absenteeism research that we have conducted.

A part of the methodology and the example introduced in this paper has been submitted as a journal paper, titled "Improving Realism of an Agent-Based Model for Construction Organizational Research Using Survey Data," to the ASCE Journal of Construction Engineering and Management.

THEORETICAL BACKGROUNDS

Axtell and Epstein (1994) argued that an agent-based model's performance can be evaluated by what empirical standards the model meets. They summarized four different performance levels of an agent-based model based on this idea. The four levels are as below (in each level, macro-structure refers to the system-level behavior of simulated organization while micro-structure refers to the individual agent's behavior):

- Level 0: "The agent behavior rule is in qualitative agreement with the behavior of real people."
- Level 1: "The model behavior is in qualitative agreement with empirical macro-structures."
- Level 2: "The model behavior is in quantitative agreement with empirical macro-structures."
- Level 3: "The model behavior is in quantitative agreement with empirical micro-structures."

Although the target performance level may vary according to the modeling objective and the availability of necessary empirical data, agent-based models used for engineering or management research would be required to be realistic, which means that the model behavior has to be in quantitative agreement with empirical data to some extent (i.e., Level 2 or Level 3). This is because simulation models in this area are often developed with an ultimate goal to provide pragmatic assistance in decision making to deal with a current phenomenon in an organization. Therefore, researchers who use ABM for engineering or management research should have an idea regarding how the necessary data can be collected and appropriately used to show a quantitative agreement of the model behavior with empirical data.

In a broad sense, data collection methods can be divided into observations and interviews (including questionnaires). An advantages of an observation is that a researcher can directly measure the variables of interest, and thus does not need to worry about the reliability of the collected data, while the disadvantages of an observation is that the observer affects the situation and that observation usually requires large efforts and time (Robson 2002). On the other hand, interviewing requires relatively less effort. However, developing questionnaires and administrating surveys requires considerable skill and experience, and making interviewees (i.e., respondents) cooperate in the survey is a very important issue for obtaining quality data (Robson 2002).

Arguably, observation is not the most viable method to collect psychological properties such as perception and attitudes (Robson 2002). Therefore, surveys are often used to assess such psychological properties. Psychologists and other social scientists have developed measuring instruments to assess people's properties that are difficult to observe, and these measuring instruments typically use a measurement scale (e.g., the Likert scale) (Robson 2002).

Therefore, if a researcher uses surveys to collect perceptual, attitudinal, and/or behavioral data to use the data as a comparative basis for simulation model validation, an important issue is to make the survey data and the simulation data comparable to each other. If the type and scale of variable is the same between the variable used in the simulation and the variable used in the survey, this issue does not exist. Observable and measurable variables (e.g., absence rate) can be designed as a model variable and as a survey variable in the same type with the same scale. However, oftentimes perceptual and attitudinal variables are neither observable nor measurable, thus discrepancy between the model variable and the survey variable is unavoidable. Attitudinal or behavioral variables in human behavior studies are most commonly measured using a scale, such as a 5-point or 7-point Likert scale, whereas the attitudinal or behavioral variables used in an ABM simulation model may have a specific operational definition. An example of this is that the level of others' social influences can be modeled as a dimensionless ratio ranging from 0 to 1 in the simulation, while the most viable way to measure this in reality is taking people's responses to the statements like, "I care about my peers' opinions when I make a day-to-day decision" on a scale of 1 (strongly disagree) to 5 (strongly agree). Therefore, the data of different types need to be transformed to be compared to each other.

PROPOSED METHODOLOGY

As previously mentioned, agent-based models used for engineering/management research often need to be a Level 2 or above model (i.e., quantitative agreement of the model behavior with empirical data) due to the pragmatic modeling objectives. However, a Level 3 model is extremely difficult to achieve in organizational research because the microscopic agent behavior should be in quantitative agreement with real individuals' behavior. Therefore, the proposed methodology in this paper aims at an agent-based model that meets the Level 2 criteria (i.e., macro-level, quantitative agreement). In other words, the proposed methodology addresses how to conduct a test to ensure the macro-level quantitative agreement of the model behavior with empirical data.

Figure 1 shows the conceptual flow diagram of the proposed methodology for enhancing the credibility of model behavior by comparing simulation data with survey data. It begins with modeling simulation variables and survey variables. As a first step, the constructs of interest are transformed as agent-based model variables through operationalization (A in Figure 1). A computational model usually consists of a set of variables that represent the states of an agent and a set of computational rules for the processes by which the variables change over time (Harrison et al. 2007). Once a simulation model is created with all of the variables and computational rules, simulation input data can be generated by a random number generator following a distribution for each input variable, and then simulation output data is generated as a simulation runs (C in Figure 1). Synchronously, a researcher can develop a set of questions as a measuring instrument of the constructs of interest (B in Figure 1), and then collect data using the questionnaire (D in Figure1).

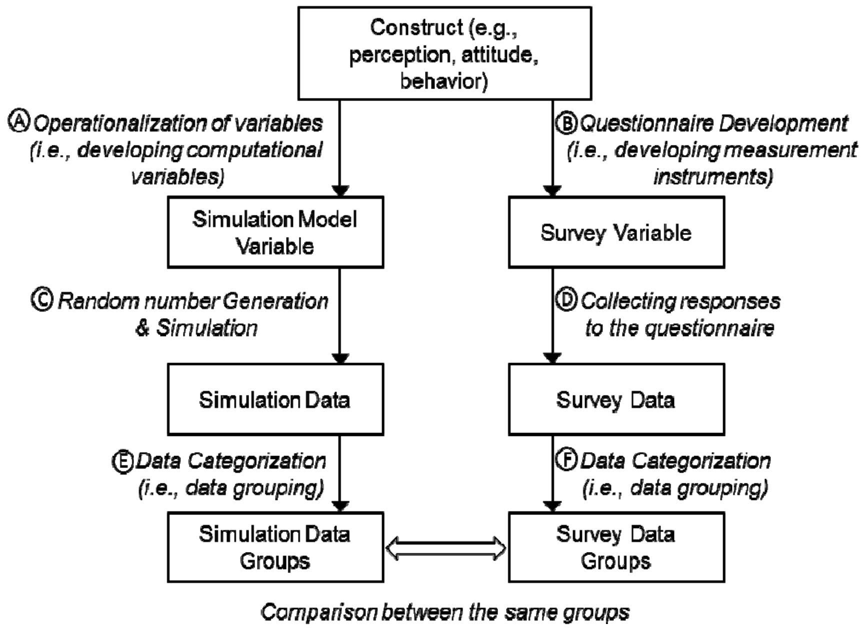

Figure 1. Conceptual flow diagram of proposed methodology

Next, the simulation data and the survey data are transformed into categorical data (E and F in Figure 1). This step is required when the type and/or scale of variables are different between the simulation data and the survey data, as previously discussed. The categorization of data is therefore an effort to reconcile the discrepancy between the simulation model variables and the survey variables. There can be several ways to categorize the data. For example, a modeler can apply predetermined boundaries for categories and rearrange the data accordingly. Alternatively, a modeler can use statistical boundaries such as 1st quartile, 2nd quartile (median), and 3rd quartile as a boundary of each category. Additionally, a modeler might consider machine learning algorithms for automatically classifying the data into several categories. Once a categorization is complete, a modeler can compare the model output with the corresponding real data for each category (i.e., group) to validate the model behavior.

AN EXAMPLE: WORKER ABSENTEEISM STUDY

In this section, an illustration of the proposed methodology is presented using the construction worker absenteeism study we have been conducting. The objective of this study has been to understand the group-level absence patterns that emerge from individuals' behavior and interactions. More specifically, we have attempted to understand the impact of construction workers' characteristics concerning adaptation and self-regulation on the group level (i.e., crew level) absence rate over time.

Therefore, the three constructs—the level of social adaptation, the level of formal rule adaptation, and the level of strictness in self-regulation—have been operationalized for the simulation model (Ahn et al. 2013a) and the indicators of the constructs have been developed as measuring instruments (Ahn et al. 2013b). Table 1 shows the constructs, the simulation variables, and the indicators used in the survey.

Table 1. Examples of simulation variables and survey variables used in the research

Construct	Operational definition of variables used in the simulation	Indicators used in the survey (Instrument)
Formal rule adaptation	The weight a worker gives to the formal absence standard over the perceived social absence norms	Anxiety about breaking formal rules ("I worry about breaking project rules")
Social adaptation	The inverse of the amount of days a worker would take to fully accept others' behavior as a norm	Perceived salience of social norm ("My team members have a high degree of agreement about which behavior is wrong")
Strictness in self-regulation	The lowness of the probability that a worker is absent when the personal absence record is over the internal absence standard	Desire to outperform ("I would like to perform a better job than the average worker in the team")

In this research, we developed an agent-based simulation model and developed a survey instrument in parallel. After the agent-based model was created, the model was specified with different settings for the three variables (i.e., agents representing workers in the simulation were characterized by different levels of formal rule adaptation, social adaptation, and strictness in self-regulation), and simulations were performed. A simulation with each setting was repeated 100 times, and the model output (i.e., group absence rate) was averaged to get a representative model behavior with each setting.

In the meantime, we collected data from a total of 228 construction workers of 29 crews. The construction workers were recruited from three different job sites in Ann Arbor, Michigan. We have grouped the 29 crews into 4 groups according to the level of social adaptation (i.e., "low" and "high" groups) and the level of formal rule adaptation (i.e., "low" and "high" groups) (i.e., categorization). The level of strictness in self-regulation was not used for categorization, because the variable did not have enough distinguishing power within our dataset. The "low" and "high" was determined by whether the crew average absence rate is larger or smaller than the average of the entirety of workers participating in our survey. Each group is denoted as "LS-LF", "HS-LF", "LS-HF", and "HS-HF" ("L" means low, "H" means high, "S" means social adaptation, and "F" means formal rule adaptation). Table 2 shows the mean and the median of each group's absence rate (day/year).

Table 2. Crew average absence rate of each group from survey data

Group	LS-LF	HS-LF	LS-HF	HS-HF
Median (day/year)	4.33	3.18	3.16	1.87
Mean (day/year)	4.8	5.24	3.29	1.93

We have examined our agent-based model's behavior by comparing it with empirical data. To do this, we have found the location of each actual group on the simulation results based on the level of social adaptation and the level of formal rule adaptation. Figure 2 is a graph of the simulation results with the corresponding points of actual groups marked on it. It is shown that the average absence rate from the simulation and the average absence rate from the survey are comparable for each group (i.e., "LS-LF", "HS-LF", "LS-HF", and "HS-HF"). With this result, the confidence with the performance of the agent-based model has been improved.

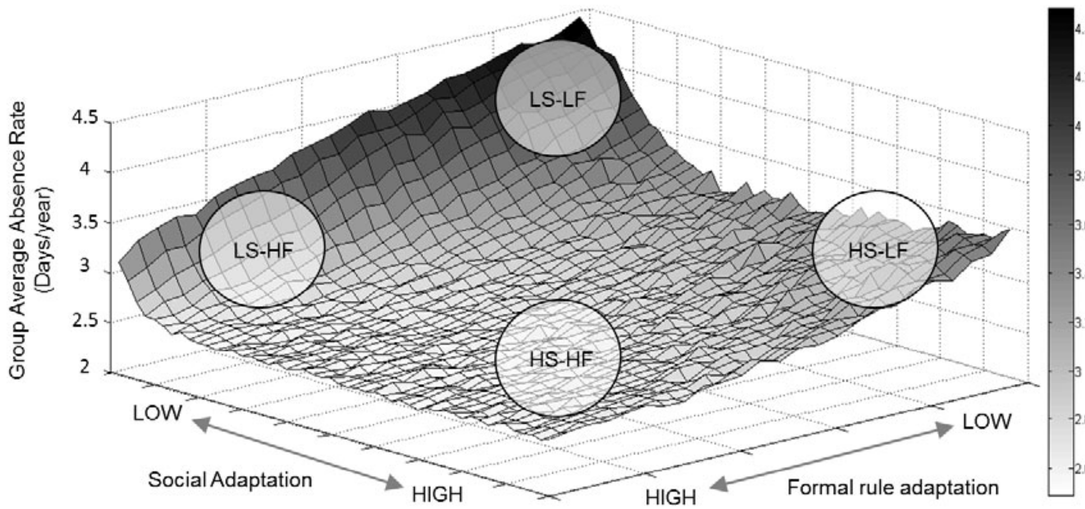

Figure 2. Simulation results with the corresponding points of actual groups marked on it.

CONCLUSIONS

The objective of this paper is to propose a methodology that can help enhance the credibility of an agent-based model for the study of workers' social behavior by comparing simulation data with survey data. It is suggested that the different types of variables measured by different scales can be compared using a data categorization method. Although detailed information in simulation data or in survey data is missed when raw data is transformed into categorical data, demonstrating the correspondence of simulation data to survey data using a data categorization method can be an effective means to enhance the credibility of an agent-based model.

This paper has limitations. More data need to be collected and analyzed. In this study, we interviewed 228 construction workers, but when the data was aggregated into crew level, the 29 data points were used in the group-level analysis. Also, different data categorization methods need to be considered. These topics will be addressed in our future research.

ACKNOWLEDGEMENT

The work presented in this paper has been financially supported by a National Science Foundation Award (SES 1127570).

REFERENCES

Ahn, S., Lee, S. and Steel, R. (2013a). "Effects of Workers' Social Learning: Focusing on Absence Behavior." *J. Constr. Eng. Manage.*, 139(8):1015–1025.

Ahn, S., Lee, S. and Steel, R. (2013b). "Construction Workers' Perceptions and Attitudes Toward Social Norms as Predictors of Their Absence Behavior." *J. Constr. Eng. Manage.*, Submitted.

Ashforth, B. E., & Mael, F. (1989). "Social identity theory and the organization." *Academy of management review*, *14*(1), 20-39.

Axtell, R. L., & Epstein, J. M. (1994). "Agent-based modeling: understanding our creations." *The Bulletin of the Santa Fe Institute*, *9*(2), 28-32.

Bandura, A. (1991). "Social cognitive theory of self-regulation." *Organizational Behavior and Human Decision Processes*, 50, 248–287.

Choudhry, R. M., & Fang, D. (2008). "Why operatives engage in unsafe work behavior: Investigating factors on construction sites." *Safety Science*, 46(4), 566-584.

Chi, S., Han, S., & Kim, D. Y. (2012). "The Relationship between Unsafe Working Conditions and Workers' Behavior and Their Impacts on Injury Severity in the US Construction Industry." *Journal of Construction Engineering and Management*, 139(7), 826–838.

Friedkin, N. E. (2004). "Social cohesion." *Annual Review of Sociology*, 409–425.

Heinrich, H. W., Petersen, D., and Roos, N. (1980). *Industrial accident prevention*, 5th Ed., McGraw-Hill, New York.

Hendrickson, C. (2000). *Project Management for Construction– Fundamental Concepts for Owners, Engineers, Architects and Builders*, 2nd Ed., available at http://www.ce.cmu.edu/pmbook/, accessed on 7 October 2013.

Hinze, J., Huang, X., and Terry, L. (2005)."The nature of struck-by accidents." *J. Constr. Eng. Manage.*, 131(2), 262–268

Maloney, W. F. (1983). Productivity improvement: The influence of labor. *Journal of construction Engineering and Management*, 109(3), 321-334.

Mitropoulos, P., & Memarian, B. (2012). "Team Processes and Safety of Workers: Cognitive, Affective, and Behavioral Processes of Construction Crews." *Journal of Construction Engineering and Management*, 138(10), 1181-1191.

Mohamed, S. (2002). "Safety climate in construction site environments." *Journal of construction engineering and management*, 128(5), 375-384.

Robson, C. (2002). *Real world research: A resource for social scientists and practitioner-researchers* (Vol. 2). Oxford: Blackwell.