National Taxpayer Advocate

Simulating EITC Filing Behaviors: Validating Agent Based Simulation for IRS Analyses: The 2004 Hartford Case Study

2007

Simulating EITC Filing Behaviors: Validating Agent Based Simulation for IRS Analyses: The 2004 Hartford Case Study

Prepared for: The National Taxpayer Advocate
Internal Revenue Service
1111 Constitution Avenue
Washington D.C. 20224-0002

Prepared by: Professor Kathleen M. Carley, Director, CASOS
Institute for Software Research
Carnegie Mellon University
Dr. Daniel T. Maxwell, Senior Principal, IDI
Support provided by CASOS staff – Jessica McGillan, Mike Kowalchuk

1 September, 2007

Executive Summary

The IRS Office of Program Evaluation and Risk Analysis (OPERA) initiated a research program in 2004 to explore the feasibility of using multi-agent (agent based) simulations to help inform decision making about how to more effectively and efficiently administer the US tax system. This program is being conducted under the sponsorship of the National Taxpayer Advocate, and is part of a research effort originally initiated in 2003 at the request of the National Taxpayer Advocate to explore the applicability of a variety of technologies to understanding the dissemination of abusive tax schemes. The Office of the National Taxpayer Advocate expressed interest last year in exploring the applicability of this type of simulation for supporting other analyses. The National Taxpayer Advocate's interest centered on increasing confidence in the validity of the simulation for representing subsets of the population that are of high interest to the Taxpayer Advocate. To accomplish this, the National Taxpayer Advocate requested that the simulation team attempt to replicate an actual tax related event that occurred in the US population. The specific request was to recreate the IRS EITC Certification study experience in Hartford County, Connecticut, for tax year 2004.

The project, executed by the research team at Carnegie Mellon University, achieved the following three goals:

Goal 1: Demonstrated the ability to represent the diverse types of events that occur in a complex social environment in the Construct simulation.

Goal 2: Demonstrated the ability to "tune" the Construct simulation to reasonably approximate a real world experience.

Goal 3: Examined via the Construct multi-agent simulation the relative impact of the diverse events that actually occurred as compared to other sequences of events that might have occurred (counterfactuals).

Demonstrating achievement of the first goal, the simulation was able to represent a set of key events and agent behaviors that provided insight into the filing behavior of the target population. That said, the insights were more immediately visible to the simulation development team. Some key observations relating to this goal are that nontrivial simulation development was required to provide the functionality necessary to represent this scenario. Taxpayer agents needed to be parameterized differently for this effort than for previous phases. The lessons of this experience are informing the evolving design of the taxpayer agents and the simulation overall. The representation of targeted notices sent to taxpayer agents meeting specific demographic criteria identified necessary functionality for understanding how changes to IRS notice programs will affect taxpayer behavior. Implementing this very precise type of communication into a multi-agent simulation was a nontrivial challenge. This required the pursuit of some advances to the state of the art in multi-agent simulation and consumed significant resources. That said, this advanced functionality was successfully implemented and is available for use in future Virtual Experiments.

The second goal to tune the simulation to appropriately match the Hartford experience emphasized comparative evaluation of the event sequences, population description and behaviors, and event timing. All three of these dimensions were successfully tuned, or matched to the Harford experience. The accompanying table illustrates the match that was achieved with the EITC Study population as reflected in the 2007 Study.

Characteristic	EITC Study Population	Simulated Population
Percent don't file a return	0.12	0.15
Percent file a return	0.88	0.85
Percent claim EITC with children	0.58	0.62
Percent claim EITC with one child	0.31	0.35
Percent claim EITC with two child	0.27	0.27
Percent file a return and single	0.19	0.20
Percent file a return and married	0.06	0.04
Percent file a return and head of household	0.75	0.76
Percent file a return and male	0.59	0.60
Percent file a return and female	0.41	0.40
Percent file a return and under 31	0.37	0.39
Percent file a return and 31-40	0.29	0.27
Percent file a return and 41-50	0.23	0.23
Percent file a return and over 50	0.11	0.10

It is important to note that the data do not match perfectly. This is by design. It would be possible to force the simulation to match every parameter perfectly. However, that tightly controlled level of specification would then constrain the ability of the simulation to naturally change in response to changes in the independent variable(s), rendering it invalid for use in Virtual Experiments.

The third goal, to examine via a virtual experiment the relative impact of diverse real-world events on the multi-agent simulation, demonstrated that the simulation was capable of consistently generating results that closely approximated many of the response measures across several possible situations.

Introduction

The IRS Office of Program Evaluation and Risk Analysis (OPERA) initiated a research program in 2004 to explore the feasibility of using multi-agent (agent based) simulations to help inform decision making about how to more effectively and efficiently administer the U.S. tax system. The research effort is focusing on enhancing the functionality of an existing simulation, called Construct, to represent explicitly the key tax related beliefs, knowledge, decisions, and behaviors of taxpayers; as well as the diffusion of tax related information around a city sized population. The initial simulation development efforts provided encouraging results that were reported to the Service research community in June of 2006. (Carley & Maxwell, 2006)

The Office of the National Taxpayer Advocate is sponsoring this research. The National Taxpayer Advocate expressed interest in increasing confidence in the validity of the simulation for representing subsets of the population that are of high interest to the Taxpayer Advocate. To accomplish this, the National Taxpayer Advocate requested that the simulation team attempt to replicate an actual tax related event that occurred in the US population. The specific request was to recreate the IRS EITC Certification study experience in Hartford County Connecticut for tax year 2004.

The project had three major goals. Specifically:

Goal 1: Demonstrate the ability to represent the diverse types of events that occur in a complex social environment in the Construct simulation.

Goal 2: Demonstrate the ability to "tune" the Construct simulation to reasonably approximate a real world experience.

Goal 3: Examine via the Construct multi-agent simulation the relative impact of the diverse events that actually occurred as compared to other sequences of event that might have occurred (counterfactuals).

These goals were accomplished over a period of approximately 120 days by the research team at Carnegie Mellon University's Center for the Analysis of Social and Organizational Systems (CASOS). Some additional simulation functionality was added to the simulation to represent the specific filing decisions associated with EITC and with the behaviors of opinion leaders in a community.

The effort and its results are described in the sections that follow. The next section provides a short description of the Hartford scenario, with emphasis on the factors that were most relevant for informing the simulation effort. We then describe the key characteristics of the Construct simulation, highlighting the behaviors that are relevant to simulating the key events in Hartford. This section is responsive to Goal 1 above. After the scenario and the simulation are described, we report the results of the successful simulation tuning effort; documenting achievement of Goal 2. The next section describes a demonstrative "Virtual Experiment" that was conducted using the simulation. This experiment explores

counterfactual situations, or what might have happened if the events in Hartford had unfolded differently, satisfying Goal 3. Finally, we will conclude with a set of conclusions and recommendations that address simulation development and the role multi-agent simulation might play in supporting Service decision making.

The Hartford Scenario

The IRS has a long term ongoing and multi-faceted program initiative that is attempting to ensure fairness and compliance in the administration of the Earned Income Tax Credit (EITC). One of the initiative's goals is to reduce EITC over-claims without adversely affecting participation among eligible taxpayers. The tax year 2004 report in support of this initiative documents a qualifying child certification test that was conducted in Hartford County Connecticut between November 2004 and December 2005. (IRS, 2007).

The scenario began unfolding when information of the test reached the mayor of Hartford City in early November 2004; that information triggered a series of events that influenced the certification and filing behavior of the city's (and likely some county) residents. For purposes of our analysis these events are information that is communicated to the potential EITC population that influences their decisions. There are five key "information" events in the simulated scenario:

- Hartford City mayor announces IRS agrees to three-week delay in targeted IRS review of Hartford taxpayers.
- November 28, 2004 – IRS sends first notice to selected taxpayers;
- November 29, 2004 – Hartford City sues IRS;
- January 5, 2005 – IRS reminder notice sent to those that did not certify;
- January 31, 2005 – Hartford City mayor launches EITC tax preparation program.

These events are analytically interesting because they represent a set of mixed messages; some encouraging taxpayers to comply with the program and some discouraging compliance. They become even more interesting because both sets originate from authoritative sources (called opinion leaders in the simulation) and the mayor's message changes tone over time.

The test group consisted of 8041 taxpayers who were EITC claimants in Hartford County by April 15, 2004, and who claimed qualifying children in 2003 that could not be verified. This part of the population is identified in figure 1 inside the smallest ellipse. There are two other segments of the population that are relevant to our analysis. First, there are 47,000 taxpayers in Hartford County that are similar to the test group, with the exception that in the prior year they did not claim qualifying children, or the eligibility of the claimed children was verified by the IRS. The outer ellipse completes the county population identifying the high income taxpayers. The focus of the analysis are the regions of the population that includes the ellipses showing the general EITC population and

the 8,041 taxpayers. Another analytically interesting feature of the Hartford County population is the distribution of taxpayers between the city of Hartford and the remainder of the county. This is represented by the horizontal ellipse. This geographic distribution is especially important for two reasons. First, the initial opposition to the certification test originated with the mayor of the city. So he is a formal opinion leader for only a subset of the population. Second, the density of EITC claimants as a proportion of the total population would be expected to be higher in the city compared to the remainder of the county based on the demographic data that is available.

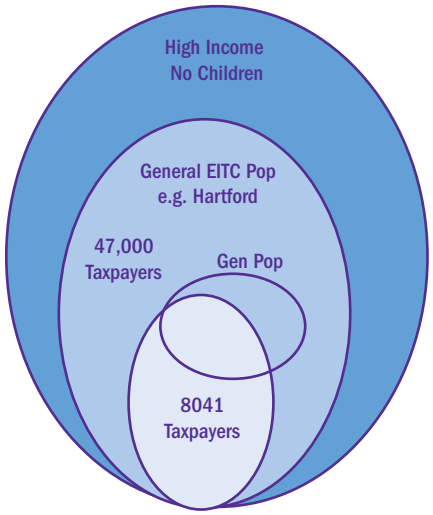

Figure 1: Hartford Population Description

Within the population of interest it is important to gain a more highly resolved understanding of the population. This detail is what allows analysts and simulation developers to increase the fidelity of a simulation. More importantly this increased resolution better matches the types of decisions about how to support and respond to diverse populations Service executives face in administering the tax system. In our case the data from the FY 05 report indicates that our population of interest has the following key characteristics:

- Filing Status (Married – three percent, Female – 37 percent, Male – 60 percent);
- Income (0-15K — 42 percent, 15-30K — 38 percent, 30-35K — 20 percent);
- Age (0-29 — 34 percent, 30-39 — 29 percent, 40-49 — 23 percent, 50+ — 14 percent);
- Paid preparer use (72 percent of all taxpayers without children, 76 percent of EITC with qualifying children eligible taxpayers).

Simulation Description

CASOS' Construct simulation has been richly developed and applied to numerous social science related research projects and governmental analyses over the past decade. A complete description of the simulation is beyond the scope of this effort, but a very complete reference is Schreiber & Carley (2004). Efforts to effectively represent taxpayer (agent) behaviors and collect outcome data that are relevant to tax administration have been underway for some time. So, there is an existing set of relevant behaviors and measures of effectiveness that provided a foundation for this effort. These are described in Carley & Maxwell (2006) and the interested reader is referred to that source for more complete information.

Multi-agent simulations consist of a few key parts that we will describe briefly here in the context of the Harford test. First, there is a population of agents that interact with each other in a simulated environment. Second, each agent possesses characteristics, or attributes, that give it an identity in the population of agents. Third, the agents have a set of beliefs, knowledge, and behaviors that reflects how they interact with other agents, as well as key actions they will perform. The set of behaviors available to an agent is a function of its attributes. And finally, there is a set of outcome measures that are collected reflecting the status of key variables in the simulation. (e.g. number of rejected EITC claims).

The scenario being modeled runs over a period of one year. It begins immediately following the April 15th filing deadline in 2004 and runs through the Tax year 2004 filing deadline April 15th 2005. The simulation is a time stepped model, with each time step representing one week on the calendar.

The Construct **taxpayer agent population** for this effort consisted of 3218 agents that represent the EITC population of Hartford County. Fifty percent of the agents (1609) were assigned attributes consistent with the study population, and the other half represented the remainder of Hartford's General EITC population. This distribution of agents that emphasizes the study population as a percentage of the total implements a design technique called matched sampling (Rubin, 2006). We will assume for the purposes of this analysis that all of the agents "know of" the existence of EITC.

There are also two special types of agents in this scenario. The **Mayor of Hartford** is represented explicitly as an agent that broadcasts messages to multiple agents. And, the **IRS notices** to the study group were also represented as a special type of agent that interacts with taxpayer agents meeting specific criteria, and a history of the interactions is maintained. Implementing the IRS notices required some significant enhancements to the simulation software. This was necessary to ensure that initial notices were sent with certainty to the correct subset of the taxpayer agent population. Still more enhancement to the software was required to ensure that follow up notices were sent with certainty only to taxpayer agents that received an initial notice and did not respond to the first notice.

The taxpayer agents are described by six attributes that were identified by IRS experts as relevant to understanding the behavior of the general EITC population, as well as informative for cross classifying the agents inside the broader population. These are:

- Filing Status – Married, Unmarried Female, Unmarried Male;
- Preparation – Unpaid or Paid;
- Age – 0-29, 30-39, 40-49, 50+;
- Income – 0-15,000, 15,001-30,000, 30,001-36,000;
- Children – 0,1,2+; and
- Locale – Hartford County, Hartford City.

These characteristics are assigned to each agent so that the population is consistent with the demographics as they are described by census data, and the parameters of the sample matching process. The preparation data was provided by the EITC 2005 study, with 72 percent of the population without children and 76 percent of those with children using a paid preparer.

TABLE 1, Relevant Facts and Beliefs

Facts Associated	Category
1	Know of EITC (Knowledge)
10	How to claim EITC (Knowledge)
12	How to certify for EITC (Knowledge)
1	Know of certification process (Knowledge)
6 (3 yes 3 no)	Eligibility belief (Belief)
7 (4 yes 3 no)	Certification belief (Belief)

All of the agents in the simulation possess facts and beliefs. These are a binary representation where a 0 reflects the absence of a belief or fact and a 1 indicates it is present. As concepts get more complicated they possess more potential facts or beliefs. Table 1 illustrates the key facts and beliefs that are relevant for this scenario.

A simulation run extends over 52 weeks, with each simulation time period equal to one week. There are two types of events; external events that are scripted into the scenario, and taxpayer events that potentially occur at each time step based on the agents' current knowledge and beliefs. There are five possible external events in the simulation occurring at specified times. The calendar dates, event, and simulation time period are identified below.

- November 8, 2004 – First anti IRS message from the mayor – time-period 29;
- November 28, 2004 – Initial IRS soft notice sent – time-period 32;
- November 29, 2004 – Second anti IRS message from the mayor – time-period 32;
- January 5, 2005 – IRS reminder notice – time-period 37;
- January 31, 2005 – First pro IRS message from the mayor – time-period 41.

In addition to the scripted events, the agents have a set of behaviors that may (or may not) occur at each time step. There is a base set of behaviors where they communicate with other agents, as described in the cited Construct foundation literature. The set of custom behaviors that was designed and implemented for the EITC project follows:

- Receive first IRS soft notice;
- Receive second IRS soft notice;
- Listen to Mayor;

- Decide to get certified;
- Get certified;
- File & claim EITC; and
- File & don't claim EITC.

Each of these behaviors has a set of unique characteristics that are worthy of exploring in a little more detail.

Taxpayer notices from the IRS are sent with certainty to the specified study population. Once they get the notice, it is read with a very small probability. When the notice is read the agent can randomly learn up to fifty percent of the content. If the notice is not read, the likelihood that it will be read in subsequent time steps decays exponentially. Notices can be read multiple times, with the exponential decay function reinitializing each time it is read. This simulation functionality was tested, and showed significant impact on the counterfactual scenarios in the virtual experiment.

The messages from the Mayor are communicated to multiple agents. The behavior of any one agent will then depend on which messages are received from the mayor, what other information the agent has received, and the rest of its descriptive variables.

The decision to certify is influenced by a set of factors that is defined by the agents' initial beliefs and the information they receive over the course of the simulation scenario. The specific factors involved are:

- They are subject to social influence – others tell them to or not to certify;
- They believe they are eligible;
- They believe the IRS will freeze their refund and force them to prove eligibility; and
- They have information about certification.

Once the decision is made to certify, the taxpaying agent is successful 75 percent of the time in receiving certification. This event is random and at a rate that is consistent with the data provided in the 2005 study.

The two potential filing decisions identified imply a third possibility. The agent might choose not to file. It turns out that this behavior choice for this subset of the taxpaying population is over ten percent. The decision not to file could be because the taxpayer had a change in status that eliminated the need to file (a compliant decision), or the taxpayer could be choosing to become noncompliant.

There are two different sets of behaviors and decision criteria in the simulation associated with claiming EITC when filing: self prepared and paid preparer. In order for self preparer agents to claim EITC on their returns conditions 1 and 2 must be met, and either 3 or 4 or 5 must be met.

1. **Know of the EITC**
   - Mark 1 fact as representing that the credit exists
   - Agent knows that fact
2. **Have Sufficient How To Facts to participate in EITC**
   - Agent knows 50 percent of the how to facts associated with the credit
   - Complex - credit is 12 facts
3. **Random decision to participate**
   - Ultimately to be related to risk taking behavior
   - Ultimately to be related to other psycho-socio factors
4. **Believe they should claim the EITC**
   - Based on 6 makes sense facts – 3 suggest to engage, 3 to not engage
   - Belief > threshold for engagement
5. **Believe they are eligible for EITC**

If the agent uses a paid preparer the simulation logic for claiming EITC is as follows:

- Claims the EITC for the agent
- Knows of EITC and claims it
- If agents think they are eligible then so does the preparer
- Mis-claims are based on the 2005 study data

Simulation Calibration

The process of tuning a multi-agent simulation model, also called calibration, has a long tradition in the modeling community. Model tuning is the process of adjusting a computational model to reflect the features of empirical data (Carley 1996). The empirical data is historic in nature and provides the foundation for a case study comparing and contrasting what was observed with the simulation results. The result of tuning is a set of model parameters that best represent one instance of the phenomena of interest. This provides a conceptual reference point for validation, analysis, and interpretation of the computational model. Successfully tuning a model demonstrates that it is sufficiently rich to capture the behavior observed in the past, and that it is likely to be sufficiently accurate that it can be used to draw conclusions about similar behavior in the future assuming that there are no major changes in the environment (*e.g.,* change to a Value Added Tax system).

Tuning or calibration is the second in a series of validation steps going from face validity (it feels right) to accurate predictions (x number of people will file with this error in this city in this year), that are consistent with all of the uncertainties that might influence the

outcomes of interest. The issue with respect to validation is not whether a model is valid, but to what degree.

Model validation should always be done with an analytic purpose in mind. Although models are routinely criticized for lack of realism, in point of fact the level of realism and therefore the associated level of validation should be chosen to match the intended purpose of the model. For most purposes, particularly when the use of the model is to think through the basic policy issues, tuning to a level sufficient for generating credible interval measures that compare and contrast the independent variables in a Virtual Experiment (VE) is consistent with the purpose of the model.

The process of tuning a model involves data analysis of virtual and real data, parameter adjustment, repeated execution of virtual experiments, and sometimes, simulation software development. For this study we had three key types of real empirical data that the model was "tuned" to: 1) event sequences such as when opinion leaders did what; 2) population level results on the percentage of agents (taxpayers) that claimed EITC with children in the study and control group; and 3) the timing of filing over the calendar year. Each of these data required a different set of procedures for tuning.

Sequence data. To tune to sequence data we verified that the events in the sequence could be represented in the Construct simulator, and that the ordering of the modeled events matched the real world. The result was that the timing of key opinion leader events could be specified using existing functionality and no changes needed to be made to the code to represent the ordering of the events. Small changes to simulation software were needed to represent the positive messages of the opinion leader. To test whether the events as modeled impacted the outcome, we went beyond standard tuning and ran a series of counterfactuals to examine the impact of these events in the simulation and then assessed their plausibility.

Population data. To tune based on the population data we first used generic population fractions and set up the model. Then we compared the results with the study and control groups. In general, participation was too low in the study group. This was due to the different socio-demographics in the study group. We then adjusted the agent population to match those socio-demographics. The result was that the model was able to predict EITC engagement in the study and control groups comparable to that observed. The results can be seen in a later section of this report.

Filing Timing data. To tune to this data we actually needed to create an entirely new module that enabled the simulated taxpayers to file early. Tuning did provide timing distributions qualitatively similar to the filing patterns that were observed for the EITC general population. However, the distributions are still off quantitatively. Through a series of virtual experiments, we discovered that this is due to a lack of reasoning in the agents for choosing when to file.

Table 2 highlights a comparison of the key descriptive attributes of the EITC study population and its behavior with the corresponding factor in the simulated population. The biggest inconsistency that was observed was in the percentage of the population that is claiming the EITC with one child (and so the sum with children). This inconsistency appears to be due to the different, and somewhat inconsistent, sources of data that inform the formulation of the virtual population; the combination of census and IRS data. Virtual experiments indicated that this deviation did not adversely affect the ability to comparatively analyze the different options represented in a virtual experiment.

TABLE 2, Comparison of Key Descriptive Attributes of the EITC Study Population with the Simulated Population

Characteristic	EITC Study Population	Simulated Population
Percent don't file a return	0.12	0.15
Percent file a return	0.88	0.85
Percent claim EITC with children	0.58	0.62
Percent claim EITC with one child	0.31	0.35
Percent claim EITC with two child	0.27	0.27
Percent file a return and single	0.19	0.20
Percent file a return and married	0.06	0.04
Percent file a return and head of household	0.75	0.76
Percent file a return and male	0.59	0.60
Percent file a return and female	0.41	0.40
Percent file a return and under 31	0.37	0.39
Percent file a return and 31-40	0.29	0.27
Percent file a return and 41-50	0.23	0.23
Percent file a return and over 50	0.11	0.10

Recall that the design of the virtual environment is called a matched sample, and that the actual numbers of agents in the simulated study population and control group were equal even though they were not the same size population. That said, the simulation was able to produce behaviors and results that were largely consistent with both the study group and the control group. Just as importantly the minor inconsistency in filing status that was observed between the study population and its corresponding virtual population is similarly observed between the control group and its corresponding virtual population. This is a very positive indicator with respect to the validity of the simulation for the Virtual Experiment in this demonstrative study.

EITC Virtual Experiment

The virtual experiment uses the customized multi-agent simulation to predict responses under several conditions that represent the real-world events in Hartford City, Connecticut in 2004 and 2005. The five key "information" events in the simulated scenario are:

- November 8, 2004 – Hartford City mayor announces IRS agrees to three-week delay in targeted IRS review of Hartford taxpayers;
- November 28, 2004 – IRS sends first notice;
- November 29, 2004 – Hartford City sues IRS;
- January 5, 2005 – Reminder notice sent to those that did not certify;
- January 31, 2005 – Hartford City mayor launches EITC tax preparation program.

Analyzing how well the simulation can cope with these mixed-message situations and changing influences from opinion leaders is important because it indicates the simulation's ability to handle the complex situations that arise in the real world.

Several sets of events were developed, designed to represent the events at key time-periods over the 2004-2005 year, and five replications (runs) of the simulation were applied to each:

Case 1: No intervention events occurred.

Case 2: IRS First Notice: The IRS sent out a first notice.

Case 3: IRS Reminder: The IRS sent out a first notice followed by a reminder.

Case 4: Negative Mayor: The IRS sent out a first notice followed by a reminder, and the mayor communicated a negative opinion and then followed this with a second negative opinion. The mayor's messages are negative relative to the IRS.

Case 5: Most Realistic: The IRS sent out a first notice followed by a reminder, and the mayor communicated a negative opinion followed by a switch to a positive opinion. The positive opinion message by the mayor is simply less antagonistic to the IRS than the negative message, and does encourage taxpayers to talk to tax assistance centers. This case is closest to what actually happened.

Case 6: The IRS sent out a first notice followed by a reminder, and the mayor communicated a positive opinion.

Case 7: The IRS did not send out a first notice or reminder and the mayor communicated a positive opinion.

Case 8: The IRS did not send out a first notice or reminder and the mayor communicated a negative opinion.

In general, the virtual experiment indicated that the simulated actions did not significantly affect whether people filed a return. Figure 2, below, demonstrates this result. Each bar in

the graph shows what percentage of people filed a return under each simulated condition. The last column represents the simulated conditions which most closely approximate what really happened. This corresponds to Case 5, in which the IRS sent out a first notice followed by a reminder, and the mayor communicated a negative opinion followed by a switch to a less-negative, *i.e.* a positive opinion. Additionally, for comparison, the dots in Figure 2 indicate the actual percentage of EITC study participants who filed a return. The percent who file predicted by the virtual experiment is a little low across all conditions, but differences in response rate between the conditions are not significant. (Note that the Y axis contains a range of 10 percent.) The important point here is that the Construct model is suggesting that initiating the certification process, reminding study participants, and messages sent by the mayor had little impact on who filed.

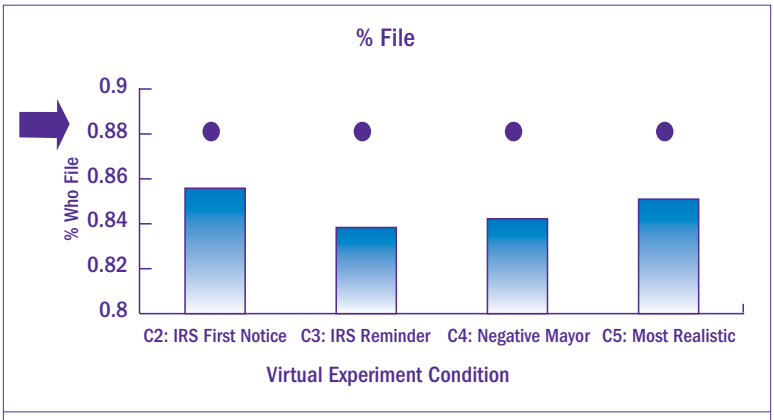

Figure 2: Comparison by condition of the response rates predicted by the virtual experiment. Dots indicate the actual filing rate.

The virtual experiment can also give insight into the possible deterrence effect of certification and the activities of the IRS and the Mayor on the response rate of the EITC study population for a particular variable. Figure 3, below, illustrates this for the percent who claimed EITC with children in the simulated and real study group.

In the actual study group there was a drop in the response rate for the percentage of taxpayers who filed and claimed the EITC with children from 2003 to 2004. This is indicated in Figure 3 in the top left. The arrow highlights this drop. This means that historically, only 58 percent of the study group claimed EITC with children after the certification process that started in the fall of 2004 began. It is important to note that a variety of factors, other than the certification process, might also have contributed to the reduction in claimants in the real world study group. Over the last few years the composition of the EITC population has changed by about a third on average, meaning that about one third stop claiming EITC and are replaced by new claimants.

We attempted to capture this general change in the composition of the EITC population in Construct. Several mechanisms were included. Some of the virtual agents had new

children; we used US birth rate. Some of the virtual children became too old to be eligible. We assumed a uniform distribution in age and simply "aged" the children, hence, approximately 1/18 of the families with children had one less eligible child assuming that they did not "gain" a child. We also simulated changes in custodianship that occurred due to divorce; for this we used average US divorce rate. While there are many additional factors that impact the composition of the EITC population, we found these three to be sufficient to account for most of the change. Other factors that might be considered in the future include: change in income due to job, college level children, and death rate.

These compositional changes are fixed across all virtual experiments. Consequently, since we simulated these types of changes to the composition of the population, variation in the results for cases 2 through 5 are due to the potential deterrence, or lack of deterrence effect due to the various interventions: notices, reminders, and mayoral messages.

In the bottom right of Figure 3 the comparison of what the Construct model predicts to the actual EITC claim rate shows the possible deterrence effect of certification under diverse interventions. The virtual experiment shows that across the different types of interventions, the response level was consistently close to the actual percentages, shown as dots. Similarly to Figure 2, there is little variation across conditions. Under most conditions, *i.e.*, most cases, the simulated results are comparable to what was observed in the real world with the study group. The critical exception here is that when only the IRS first notice is present the simulated results are lower than the real case and than other simulated conditions.

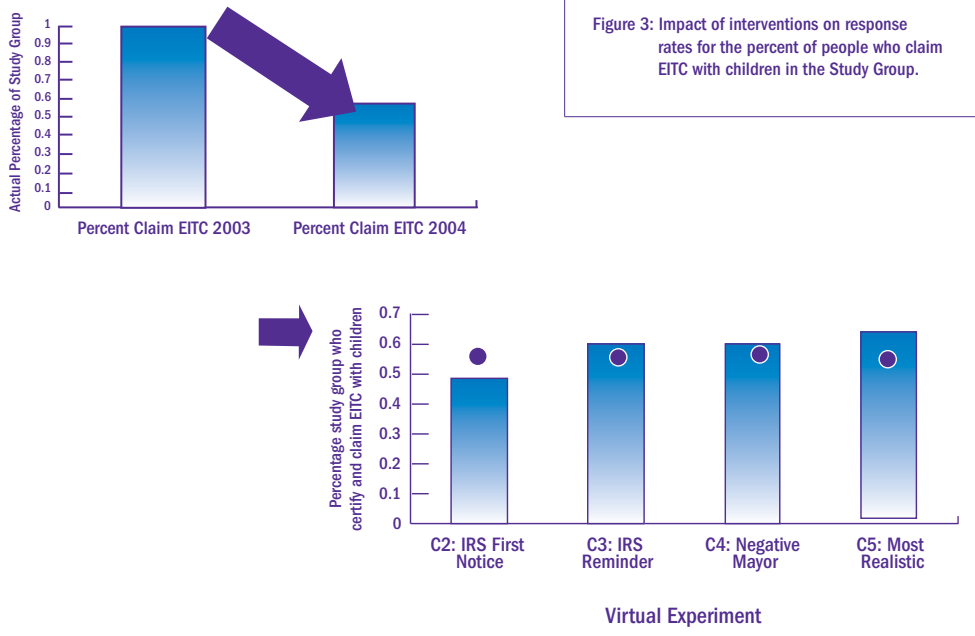

Figure 3: Impact of interventions on response rates for the percent of people who claim EITC with children in the Study Group.

Top left displays the change in percentage of filers who claim EITC with children in the actual Study Group. Bottom right displays the variation in claimants under diverse interventions. The dots on the bottom right indicate the actual percentage of filers in the Study Group who claimed the EITC with children in 2004.

It is important to note that in the real data 63 percent of the control group, as opposed to 58 percent of the study group, filed and claimed the EITC with children in 2004. This may have been due to a variety of factors, such as suppression in claims in the study group or differences in composition of the study and control group. We found that, for the simulated study group in the most realistic case, the predicted number of those who would claim the EITC with children was 62 percent. Similarly, we found that for the simulated control group in the most realistic case, the predicted number of those who would claim the EITC with children was also 62 percent. Keep in mind, the difference here is that the simulated study group had a slightly different composition than it did in the real world, and the members received both the IRS first notice and the reminder; whereas, the simulated control group did not receive the IRS first notice and the reminder. In addition, the simulation of the study group who receive only the first notice, no reminder and no messages from the mayor has a significantly lower level of filing and claiming the EITC with children. This suggests that had the IRS only sent the first notice there would have been suppression in claims; however, by sending the second notice, this suppression effect was mitigated. This also suggests that the opinions expressed by the mayor may have had little impact on claims.

Conclusions and Recommendations

The success of this effort adds to a growing confidence that multi-agent simulation could be a useful tool for informing Service analyses and decision processes. In particular, the virtual experiment emphasizes the multi-agent simulation's ability to handle a variety of possible events and produce good representations of real-world responses across those different scenarios. The ability to engage in what-if analysis is particularly informing as it brings to light the relative impact of alternative interventions both by the IRS and others.

There are a number of limitations to this study, and the results should be viewed with caution. First, many factors that played a role in Hartford were not modeled. For example, the Service met with many local groups to try and help them understand the proposed certification process. This meeting and the interaction of taxpayers with these local groups was not captured. Another key limitation is that when the mayor gave his more "positive" message, it was done in the context of encouraging taxpayers to take advantage of the tax assistance centers. The role of these centers in affecting taxpayer behavior was not modeled. A third example has to do with the timing of filing. Taxpayers tend to adjust when they file based on use of a preparer, expectation of a refund, and so on. Such timing considerations were not modeled. Consequently, the impact of intervention on when taxpayers filed and when they sought certification could not be considered.

Additional research and simulation development is required to achieve a level of maturity that is consistent with operational use of the tool. Specifically, additional virtual experiments are required to develop a body of knowledge about the simulation and its behavior, as well as sufficient confidence in its validity for representing "decision relevant" scenarios. Coincident with these validation and verification (V&V) oriented experiments additional investments in simulation functionality should be planned to respond to the limitations of the simulation that will likely be identified. Added features and sub-modules, such as ones dealing with literacy, presence of tax assistance centers, and expectations for reimbursement will clearly increase the range of policy issues the model can address. However, as features are added the time it takes to use the model to generate results for various "what-if" questions increases, and the time it takes to "retune" the model to fit historic cases, such as this EITC study, increases. Consequently, investment in features should occur with investments in parallelization, scalability studies, and continual retuning of the model as new features and modules are added.

In summary, multi-agent dynamic network models in general, and Construct in particular, can play a critical role in understanding the impact of Service activity on the taxpayer. Results have sufficient fidelity that they can support meaningful policy decisions. Initial results demonstrate that such models can have sufficient fidelity to replicate historic events and sufficient flexibility to reason about alternative histories. To move from this modest beginning to an operational tool that accurately forecasts the impact of diverse interventions on taxpayer behavior is possible; but movement should proceed cautiously considering both technological challenges (parallelization) and substantive challenges (timing of taxpayer behavior). Next steps should focus on addition of features and associated validation, consideration of other types of taxpaying behavior, and code parallelization to support higher fidelity modeling. Note, even continuing with the EITC study has benefits as it would support some validation of the impact of local groups and tax assistance centers, and possibly relative timing of tax-related activity.

References

Carley, K. (1996) Artificial Intelligence Within Sociology. *Sociological Methods & Research*, 25, 3-30.

Carley, K. & Maxwell, D. (2006) "Understanding Taxpayer Behavior and Assessing Potential IRS Interventions Using Multi-Agent Dynamic-Network Simulation", *Proceedings of the 2006 Internal Revenue Service Research Conference*, Washington D.C. June 14-15, 2006.

http://www.hartfordinfo.org/issues/wsd/government/MayorsUpdateWNTR05.pdf - *Mayor Perez Fights for Working Families*

http://www.villageforchildren.org/press/EITC percent20Kickoff percent201.28.05.pdf - *Hartford Mayor Launches Free Tax Preparation Program*

http://www.hartford.gov/news/citysuesIRS.pdf

IRS (2007) *IRS Earned Income Tax Credit (EITC) Initiative: Report on Fiscal Year 2005 Tests.* Washington D.C.

Rubin, D. (2006) *Matched Sampling for Causal Effects*, Harvard University Press, Boston.

Schreiber, C. & Carley, K. (2004) *Construct - A Multi-agent Network Model for the Co-evolution of Agents and Socio-cultural Environments*, Carnegie Mellon University, School of Computer Science, Institute for Software Research International, Technical Report CMU-ISRI-04-109.