Modeling Information Access in Construct

Brian R. Hirshman, Michael K. Martin, and Kathleen M. Carley

April 22, 2008

CMU-ISRI-08-XXX

Institute for Software Research
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract

Literacy and other interaction mechanisms have been shown to be important factors in determining whether and how individuals interact with printed and other forms of media. This lack of access to – or in cases, misinformation from – such types of media has an important social component for Construct modeling. This technical report describes the information access features in Construct, both from an algorithmic and input perspective. It describes the literacy model implemented in Construct, as well as the internet access and newspaper readership modifications to past interventions. This document is intended primarily as a reference guide for literacy and access modeling in Construct, and assumes at least some familiarity with the Construct modeling environment. A short example experiment using the literacy mechanism is described towards the end of this report.

This work was supported in part by the IRS project in Computational Modeling and the NSF IGERT in CASOS (DGE 997276). In addition support for Construct was provided in part by Office of Naval Research (N00014-06-1-0104), the National Science Foundation (SES-0452487), and the AirForce Office of Sponsored Research (MURI: Cultural Modeling of the Adversary, 600322) for research in the area of dynamic network analysis. Additional support was provided by CASOS - the center for Computational Analysis of Social and Organizational Systems 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 National Science Foundation, the Army Research Lab or the U.S. government.

Version 1.2

Keywords: construct, multi-agent simulation, dynamic network analysis, social network analysis, agents, agent modeling, literacy, information access

Contents

1 Introduction & Motivation

Construct, a multi-agent simulation platform, has sought to model media access with increasing accuracy [1, 2, 3]. For this reason, it was deemed necessary to represent the effects of literacy and information access as modifications to the system. Literacy has the first-order effect of allowing an individual to absorb information from printed material, but also has a second-order effect of allowing individuals to pass such information on to other, potentially illiterate others. Information access mechanisms, such as access to computers or newspapers, have similar first- and second-order effects. By explicitly modeling literacy and information access, Construct is now able to examine the effects of these phenomena on society.

Literacy, defined as "using printed and written information to function in society", has been an important goal of education for thousands of years [4]. While the skills of reading and writing were once held by a small band of rulers, scholars, and educated individuals, modern society has helped to provide these skills to an ever greater number of people. In the last two hundred and fifty years, literacy has spread in ways that have been unprecedented in human history [4]. Literacy allows ideas to be communicated to many people at once, as books and fliers can be printed easily to allow communication with multiple recipients simultaneously. Literacy also allows knowledge to be stored in a repository, since books and other printed materials have the potential to outlast the humans that read and write them. By modeling literacy more realistically, and by tying literacy rates more closely to socio-demographic characteristics, Construct can achieve a accurate model of information absorption from printed material.

While there are a wide variety of information sources available to members of society, the two most frequently examined sources for Construct modeling have been via print or Internet media [5]. While newspaper readership has declined somewhat over the last fifty years, Internet access became increasingly important [6, 7]. Individuals without regular subscriptions to newspapers are unlikely to read articles, let alone educational intervention advertisements which might be printed in them. Others without access to the Internet would be unable to take advantage of intervention material available on a variety of web sites. By modeling access to these interventions more explicitly, and by tying the access rates more closely to socio-demographic characteristics, Construct can better model societal patterns of information access in a manner that is more consistent with that which is seen in the real world.

Using the mechanisms recently developed, it will now be possible to distinguish between groups of agents that are literate or illiterate, those that have Internet access or lack it, and those who subscribe regularly to newspapers and those who do not. As researchers, modelers, and policymakers seek to understand the impact of such advertisements on the general population, it makes increasing sense to differentiate those who might not access or understand the material from those who do. Information may not be delivered to the some agents as easily as it may be to others; by modeling illiteracy and information access in Construct it may be possible to explore the social processes which affect these agents in particular. Using the extensions recently made to the tool, Construct's social network evolution model will now be better able to comment on this important issue.

The primary purpose of this technical report is to introduce the Construct literacy implementation and to describe its use in designing virtual experiments. To this end, the remainder of this document is organized as follows. Sections 2 and 3 describes the effects that illiteracy and access mechanisms each have on Construct agents. Section 4 describes the algorithm used for computing individual-level parameters from aggregate data. Section 5 provides further details on how to use the new measures in virtual experiments. Section 6 discusses validation, presenting literacy and information access parameters gathered about the current American population. Section 8 concludes.

2 Effects of Literacy

Literacy has three major effects in Construct. First, literacy can enable or restrict access to certain types of agents such as books, fliers, newspapers, and other printed materials. Second, since literacy is an agent attribute, literacy can slightly affect homophily calculations such that literate agents are more likely to interact with each other, illiterate agents are more likely to interact with each other, and the two groups are generally less likely to interact in the absence of other similar attributes or knowledge. Lastly, literacy can affect the information learned by agents: illiterate agents who have the ability to interact with books, fliers, or other agents requiring literacy access may receive misinformation from their (potentially mistaken) interpretation of what was communicated. Each of these effects is described in more detail in sections 2.1, 2.2, and 2.3 below.

2.1 Interaction ability

One of the most important effects of literacy is that it restricts agents from interacting with resources marked as requiring literacy. In past virtual experiments where literacy was not used, all agents were able to interact with printed materials. If attempts were made to restrict access to print media agents, it was often done by restricting the print media's interaction sphere to be a random subset of the population. Thus, the previous way to model illiteracy's prevention of access to books, fliers, pamphlets, and web pages was random with respect to socio-demographic characteristics. If two or more print interventions were active, there was no guarantee that agents who were "illiterate" and lacked access to one intervention would be "illiterate" and lack access the other. For these reasons, it was not previously possible to try and draw conclusions about the effects of illiteracy on any one particular population subset or agent group since illiteracy was randomly or inconsistently assigned.

Under the current implementation of literacy, agents who are not literate will retain access to printed material, but the information gained from this material is curtailed. These agents are able to interact with a print agent, as the agent is in the interaction sphere, but the information exchanged between the illiterate agent and the printed material is curtailed. Unlike the previous implementation of illiteracy, described above, or the implementation of information access for Internet access or readership, described in section 3, there is no direct restriction on the interaction sphere. Instead, the only factor that literacy affects is the process of information diffusion between the print agent and the illiterate agent. Thus, the number of agents who have access to the print media intervention will not change.

2.2 Homophily

Another important effect of literacy is a slight change to the homophily calculations. Literacy in Construct is represented using a per-agent parameter set to either true for literate or false for illiterate. However, parameters such as literacy are also used for homophily calculations and help define the types of agents with which an agent is similar. Agents which have similar parameter settings will have higher homophily scores, while agents with dissimilar parameter settings will have lower homophily scores. Under the previous system for literacy calculation, when interaction spheres were constricted in a random fashion, literacy had no effect on homophily. Under the new implementation for literacy, agents who are literate will be explicitly more similar to each other, agents who are illiterate will be explicitly more similar to each other, and agents who differ in terms of literacy will be less similar to each other.

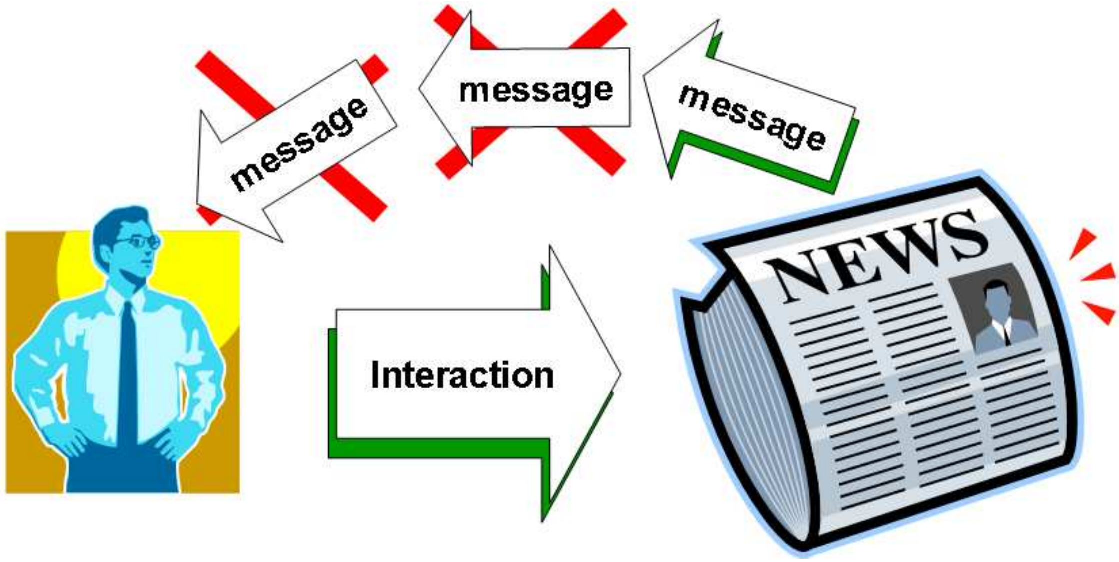
While, literacy will affect similarity by ensuring that literate agents are more likely to communicate with literate agents and illiterate agents are more likely to communicate with illiterate agents, the overall effect of this change will be rather small. Several socio-demographic parameters are often used to configure literacy and illiteracy, making literacy one parameter among many when computing agent overall socio-demographic similarity. For this reason, the overall effect of this additional homophily score will be rather small. Agents who differ on other socio-demographic parameters will still appear different even if they have the same literacy value, and while the magnitude of this difference may decrease somewhat it will still be noticeable. It will perhaps be most noticeable in helping to decrease similarity among agents who otherwise have identical socio-demographic similarity attributes. On the whole, however, literacy is just one parameter among many, and thus its effect in this respect is not particularly strong.

Like the information access parameters – described in section 3 – agents who are literate *are* affected homophily calculations. Agents who are literate are more likely to interact with other literate agents, a change which may have slight consequences for those seeking to compare old experimental results to new ones. Note that this is the only way in which literate agents are affected, as they are neither affected by the interaction ability changes described in section 2.1 or the misinformation effects of section 2.3.

2.3 Misinformation

As mentioned in section 2.1, illiterate agents have a large – but far-from-certain – chance of lacking access to print media agents. However, some agents may have access to such information, though their knowledge obtained from this information is likely to be imperfect. This distortion of information is considered misinformation, a process by which

Figure 1: Truncation of Message



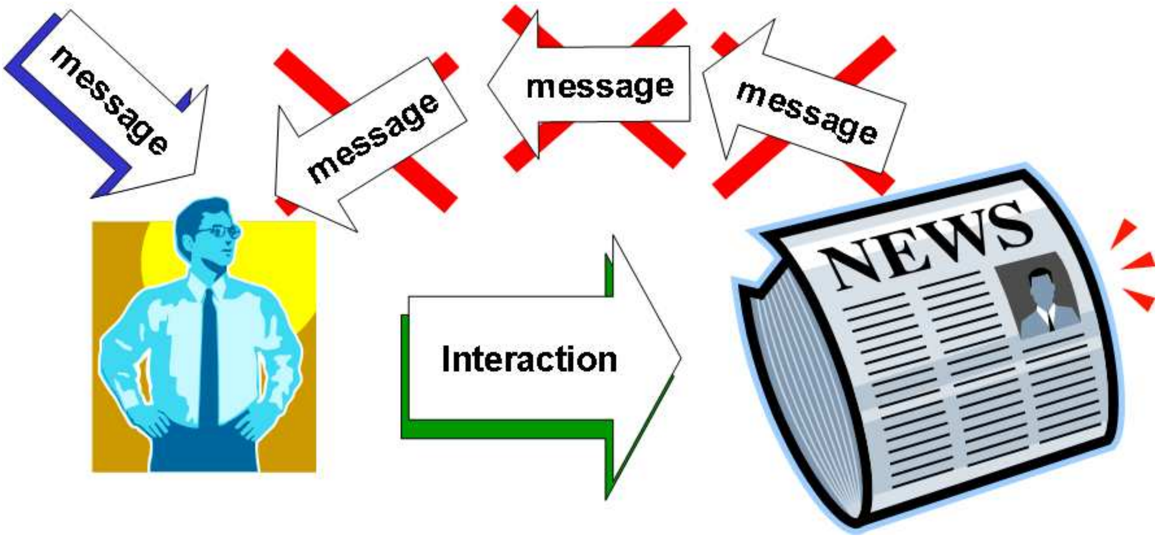
the message that the print agent seeks to send is distorted by the listener.

Agent misinformation takes three parts. The first part of the misinformation procedure is the reduction in the message size. When agents interact with printed materials, they are often able to obtain much longer messages than they are when interacting with humans. This is done because longer messages represent more intense experiences that occurs when humans interact with books as opposed to humans, but also models the fact that printed materials often contain information relevant to the virtual experiment topic. However, illiterate agents are not able to benefit from these longer messages because they cannot read the material. They may be able to gain some information from the pictures and from simple words and phrases; however, they are unlikely to absorb the entire idea that the author seeks to convey. For this reason, when a printed material agent sends a message to an illiterate agent, each element of the message sent will be eliminated from the message with a random probability based on a simulation-level parameter. If this parameter is 0.5, say, then each element of the message will have a 50% chance of being eliminated, resulting in a message that is, on average, half of the size of the origional. This procedure often results in short messages being conveyed – often no more than one or two facts – even if the sender is capable of sending a longer one. This effect is portrayed in figure 2.

The second part of the misinformation procedure is the distortion of the surviving message. Although some of the message gets through, a significant portion of that message is unintelligible to the agent. This is represented by replacing some of the selected message facts with facts randomly drawn from the suite of social facts. Each element that survived the truncation will again be compared with a simulation-level parameter; if that parameter is again 0.5, each remaining element of the message will have a fifty percent chance of being distorted. It is highly likely that such facts have nothing to do with the message, and also highly likely that the print material agent never knew such facts before sending the message. However, since the illiterate agent is unable to comprehend the true message, this distortion represents the process by which facts in the printed material are replaced partially by facts of the agent's own creation. This effect is portrayed in figure 2.

A third part of the misinformation procedure is the modification of belief facts. Some printed materials have beliefs associated with them; illiterate agents may mistake these beliefs for their opposite when attempting to interact with a particular piece of printed literature. Therefore, a certain percentage of belief facts are garbled in such a way that the opposite belief is sent to the interacting illiterate agent. If the simulation-level parameter is again 0.5, then the belief will be affected 50% of the time. Illiterate agents will then store incorrect transactive memory information about the

Figure 2: Misinformation effect



printed material agent, and will use this incorrect belief when computing the average belief in society. This effect is portrayed in figure 2, albeit indirectly: the misinformation in this step is not the relpacement with a randomly-selected message but instead is replacement with the opposite message as what is sent.

It is important to note that misinformation only applies to illiterate agents. Agents who are literate are unaffected by the misinformation procedure and will receive the full message in its entirety. They are unaffected by this procedure.

Lastly, in the current implementation of Construct, agents are marked either as fully illiterate or fully literate: illiterate agents have restricted access to printed materials and may be subject to misinformation, while literate agents are not affected by these phenomena. There is no gradation between these extremes, though future models may extend the paradigm to allow for such behavior. Importantly, the fraction of the message that is truncated, as well as the fraction of distortion of the remaining message, will be held constant for any illiterate agents. While different messages may be affected differently – some messages will be fully truncated, other messages truncated and distorted, and others completely unaffected – the parameters used for truncation and distortion are simulation-level and cannot be modified on the individual agent scale.

3 Effects of Information Access

While the literacy implementation exhibits three important effects – interaction ability, homophily, and misinformation – the other information access mechanisms only exhibit the first two effects. The interaction ability mechanism and homophily mechanisms are similar to those described for literacy, but there is no misinformation mechanism as it is assumed that those without access to the media source are not able to be misinformed via direct interaction.

3.1 Interaction ability

In earlier versions of Construct, the ability of agents to interact with different print advertisements was implemented only in a very generic manner. As described before in section 2.1, interaction spheres were cut back to represent the fact that individuals did not have access to the Internet or did not have subscriptions to newspapers. This was usually done by allowing only a percentage of the population to have access to the intervention agent, but there was no guarantee of consistency between multiple types of interventions. For instance, if there were two active web pages under the previous system, there was no guarantee that an agent with implicit "Internet access" for one of the web pages would have access to the other web page.

Under the new access mechanism recently developed in Construct, agents are given an access parameter that is specific to the intervention type. If an intervention requires a specific kind of access and the agent has that access, then the agent is able to interact with the intervention as it would any other agent. The agent will also be placed in the intervention's interaction sphere, though there is no guarantee the agent will interact with the intervention. If an intervention requires a specific kind of access and the agent does not have it, however, the agent will be unable to interact with the intervention agent. In this case, the intervention will be removed from the intervention's interaction sphere, and will also be removed from the interaction sphere of every agent that requires that particular type of access.

This, in essence, represents a communication barrier: agents without access cannot choose to initiate communication with the agent or receive information from the intervention agent. Thus, the agents will not examine the transactive memory of such interventions when scoring partners for similarity, and will ignore their beliefs when examining the generic beliefs in their interaction spheres. If an agent lacks access, however, will still be possible for that agent to learn information contained in the printed material through indirect means. Agents without access agents may be able to communicate with others who can do so, and via these intermediate agent (or chain of agents) may come to know of the material available via the print media or Internet interventions.

3.2 Homophily

As with literacy, so with access: there is a homophily component to the access mechanisms to represent the similarity between those that gain access from certain types of media outlets when compared to those who do not. In previous instantiations of Construct, in which there was no consistent implementation of access mechanisms, this similarity could not be implemented since each agent might be able to interact with one kind of intervention but not interact with another kind.

Since each of the access mechanisms is represented using an agent attribute, this attribute can be used to compute attribute similarity and affect the socio-demographic likelihood of two agents interacting. Pairs of agents who can access the Internet will be more similar to each other, as will be pairs who both cannot access the Internet; however, agent pairs in which one partner can access the Internet and the other cannot will have decreased similarities. Similarly, pairs who both have newspaper subscriptions will be more similar to each other and will be less similar to agents who do not have access to agents who lack such subscriptions. This similarity is cumulative with the fact that agents who can interact with the interventions will be more similar to each other because of their easier access to the information contained in the intervention. However, since there are a large number of other socio-demographic parameters affecting agent-agent similarity, the overall effect of the homophily change is small.

4 Inferring Literacy and Access Values

While external tools used for preparing input decks for the Construct and BioWar simulation environment often create agents from census records, these tools have no support for literacy or other access values [8, 9]. Unlike other agent attributes in Construct, then, such values have to be inferred from other sources; as described in section 6, better data is available in non-census surveys specifically designed to examine literacy. A sophisticated inference algorithm has been developed in Construct to allow for inference of individual literacy values based upon conditional literacy values as described in section 6. This algorithm relies on attributes gathered from the census data (or other data sets that are user specified) in order to determine the distribution of certain parameters in the society.

While the uses of literacy, access, and other values differ substantially in the actual Construct simulation, the preparation of per-agent literacy and access values is the same for all parameters. For ease of description, only the literacy parameter is discussed for the remainder of this section. However, Internet access, newspaper readership, and a variety of other parameters can be inferred using the same general mechanism specified below. It is important to note that the values computed for literacy will differ for those calculated for each of the information access mechanisms, though, since different data sources will likely be used.

In Construct, agents are either literate or not literate at the beginning of the experiment; agents have the same literacy value throughout the experiment. This value is determined by inferring a probability that the agent with specific characteristics would be literate, then choosing a random number to determine whether the particular instance of the agent is literate. Individual agents have multiple parameters: age, education, gender, and so forth. Some of these characteristics may be related to agent literacy (if they have are defined as described in section 5); all other parameters will be ignored. If the parameter is defined for the agent, it will be used in the calculation; if the parameter is not, the parameter will be ignored for the purposes of the calculation. Thus, the goal of this procedure is to deduce equation 1.

$$p(literacy|param_1...param_n) \tag{1}$$

In Construct experiments, there are often thousands of agents which must have literacy values assigned by equation 1. While it may be possible to do this assignment globally, assigning the literacy probabilities to all agents simultaneously, this approach was deemed too slow and un-scalable to be of practical use. Instead, a local approach was selected in which agents were assigned literacy values sequentially. Using this approach, Construct iterates through all agents who need a literacy rate, calculating a probability that the agent is literate based on its parameters as described in equation 1. This approach will result in literacy probabilities which, when aggregated, approximate the aggregate data but may not necessarily be equal.

If sufficient data were available, the per-agent literacy probability defined by equation 1 could be read off of a table. However, because literacy data is often aggregated over many characteristics, this is not often possible. For instance, the data presented in section 6 suggests that information is available for males and information is available for people in their thirties, but nothing is explicitly available for the combined group, men in their thirties. Thus, the literacy values in section 6 and the data available to most Construct researchers are generally averaged over groups in a way that makes such a table-reading approach impossible.

In order to predict the literacy value of a particular agent, then, it is necessary to infer the potential literacy rate for that agent based on the agent's group parameter values. Each of these groups are more likely or less likely to be literate; by assigning agents to these groups appropriately, it will be possible to create a distribution of agents who have the same aggregate literacies as the groups while still creating reasonably accurate individual level agent populations. The first step in following this approach is to employ Bayes rule, which can be used to rewrite expression 1 as expression 2.

$$p(literacy|param_1...param_n) = \frac{p(param_1...param_n|literacy) * p(literacy)}{p(param_1...param_n)} \tag{2}$$

Bayes rule is a mathematical expression that can be used to express a conditional probability in terms of other conditional probabilities that are easier to calculate [10]. In this equation, some of the parameters are easier to evaluate than others; however, with sufficient data and a few critical assumptions it will be possible to determine $p(literacy|param_1...param_n)$ even without exact data. The methodology used to calculate these three parameters are explained in the next three sections.

4.1 Calculating $p(literacy)$

In equation 2, the parameter $p(literacy)$, the second parameter in the numerator, is perhaps the easiest parameter to evaluate. On the one hand, it is possible to supply $p(literacy)$ directly to Construct via an input file. This explicit prior (or non-conditional) probability for literacy could then be used without additional inference. On the other hand, the more common way to create this prior probability is via inference using probabilities. Under this system, it will be possible to calculate the probability of literacy for each group times the relative preponderance of each group in society in order to determine an average overall society literacy rate. This mechanism is basically the weighted average of the literacy rates in society, weighted by each group's representative population. The equation for this calculation is displayed in equation 3.

$$p(literacy_i) = \sum_{j=1}^{M} p(literacy|param_{i,j})p(param_{i,j}) \quad (3)$$

To compute the prior value for literacy using equation 3, the input deck will be examined to determine what fraction of the agents in the deck have a particular parameter value for parameter $param_i$. This probability, $p(param_{i,j})$, will then be used as a weight on $p(literacy|param_{i,j})$, a conditional probability supplied to Construct. By summing over all values j for $param_i$, it will be possible to determine the prior literacy value for the society over parameter $param_i$. For example, when calculating the literacy probability over the parameter gender, the literacy rate for males will be multiplied by the fraction of males in society and the literacy rate for females will be multiplied by the fraction of females in society; the sum of these two groups will give an indication of the average literacy rate for this parameter.

The value $p(param_{i,j})$ can be obtained by counting all agents in the input deck which have a specific value j of $param_i$, then dividing by the total number of agents. Thus, the probability of a parameter's occurrence is considered to be exogenous and is specified by the experimenter. However, the probability is specified indirectly: the experimenter specifies the agents that comprise the society (perhaps by means of agent generators as described in other technical reports), and the literacy mechanism then performs a count over agent parameters [3].

However, since there are multiple parameters $param_i$ which can be used in the determination of literacy, it is possible that the literacy value obtained. The calculated average probability for gender, say, may be different than that for education or income. While small differences may be tolerable, this can lead to important problems with highly skewed data: for instance, consider this contrived but illustrative extended example.

- Assume that 100% of people thirty and older are literate and 0% of those over under the age are literate. Thus, those who are at least thirty are literate, and those under thirty are illiterate.
- Assume that 100% of females are literate and 0% of males are literate.
- Now assume that the experimenter decides to set up a society in which 75% of agents are thirty or over and 25% of the society is under thirty.
- Calculating the literacy rate over age color, 75% of the society should be literate: $.75 * 1.00 + .25 * 0.00 = .75$.
- However, suppose the experimenter also wants the society balanced fifty-fifty in terms of gender.
- But calculating the literacy rate over number of arms, 25% of the society should be literate: $.50 * 1.00 + .50 * 0.00 = .50$.
- This leads to an inconsistency in the overall probability rate – the society cannot be both 50% literate and 75% literate. The societal parameters are inconsistent with each other with respect to the literacy parameter.

In fact, this discrepancy is the primary reason why specifying a prior probability explicitly is discouraged: there is no guarantee that the prior probability will reflect the inferred value for the particular society, especially if the experimenter chooses to modify some of the socio-demographic parameters when creating a virtual city. While the example given above is strained and artificial, it represents an important problem that those creating virtual cities may inadvertently overlook.

There is no effective way to rectify the discrepancies in prior literacy values. Construct's method of resolving this issue is to take the average of all probabilities $p(literacy_i)$ calculated for each attribute calculated over all relevant attributes, as in equation 4.

$$p(literacy) = \frac{\sum_{i=1}^{n} p(literacy_i)}{n} \quad (4)$$

This is an unweighted average, so each parameter is counted equally. Future versions of Construct may allow for weighted averages if some parameters are determined to be more effective predictors than others; at this point, however, each parameter is assumed to be equally effective at predicting literacy rates. A warning message is provided by the literacy mechanism which will note the discrepancy, but Construct will still allow the experiment to proceed. Note that small differences may be possible – and even likely – between literacy rates calculated over different socio-demographic attributes, since individual literacy rates are unlikely to be exactly equal. This inequality is often exacerbated when small agent populations are used, or if experimenters attempt to shift parameter settings to emphasize the contribution of one socio-demographic group to the evolving society.

4.2 Calculating $p(param_1...param_n|literacy)$

The value $p(param_1...param_n|literacy)$ in equation 2 is slightly more difficult to calculate. Unlike $p(literacy)$, which can be calculated using given and inferred probabilities directly, this calculation requires several additional assumptions.

In order to obtain this expression, it is necessary to make the naive Bayes assumption, an assumption that all variables $param_1$ to $param_n$ are conditionally independent based on literacy [11, 10]. Conditional independence means that the outcome of one variable provides no further information once literacy is known: thus, it assumes that if the literacy is known than age and gender will not be related. While this assumption is known to be false – for instance, a substantial quantity of social science literature demonstrates a clear link between income and education – the naive Bayes represents the most practical way of approaching the problem of data granularity. Researchers interpreting the results from Construct should understand that the aggregate literacy statistics created by this method should still match the group statistics given as input, even though the individual-level data may be problematic as agents might be created which do not necessarily correspond to observed agents in the real world. This is a known issue and can only be resolved by obtaining more accurate and fine-grain data.

Using the naive Bayes assumption, then, this portion of the numerator can be written as equation 5.

$$p(param_1...param_n|literacy) = p(param_1|literacy) * ...p(param_n|literacy) \tag{5}$$

These parameters are more manageable, but still are not available from the given data. It is important to note that $p(param_1|literacy)$ is not the same thing as $p(literacy|param_1)$. The former predicts a parameter – say gender – based on literacy, while the latter predicts literacy based on gender. More specifically, the statement "all females can read" is not the same thing as the statement "all readers are female". While the former probabilities are available in section 6 or as inputs to Construct, the latter is not. Instead, they must be re-determined by an application of Bayes rule, as demonstrated in equation 6. Note that this equation takes the same form as equation 2.

$$p(param_i|literacy) = \frac{literacy|p(param_i) * p(param_i)}{p(literacy_i)} \tag{6}$$

In this equation, all the variables are known. The value of $p(literacy|param_i)$ is the value supplied to Construct as the specific literacy rate per parameter. The value of $p(param_i)$ is inferred from the data by counting the number of times that an agent in the input deck has parameter i with that specific value. The value $p(literacy_i)$ is the literacy value corresponding to the specific parameter from equation 3 in section 4.1; this may be different than the average literacy rate but will ensure that the value of equation 6 is less than one.

To satisfy equation 5, then, it is necessary to multiply the values calculated using equation 6 for all i parameters. This creates the desired quantity $p(param_1...param_n|literacy)$.

4.3 Calculating $p(param_1 ... param_n)$

In order to create the denominator of equation 2, it is necessary to determine $p(param_1 ... param_n)$. There are two possible ways of going about this: via direct counting and via logical inference.

The direct counting approach is perhaps the simplest to explain. The expression $p(param_1 ... param_n)$ is the result of a number of parameters logically anded together, meaning the probability of an agent having all of these parameters set. The probability of this occurring can be calculated by looking at the agents in the input deck, counting the number of agents who have that specific set of relevant parameters, then dividing by the number of agents in the population. Only parameters relevant to the literacy inference would be counted; agents who have all n parameters set but have additional parameters set (or not set) will still be counted. Under this method, the count would independent of the literacy values for the agents.

However, this approach can be problematic. If only a small number of agents are available with a certain number of parameters, then the resulting probability may be biased or even incorrect if some of the probabilities are small. For instance, consider the case where there are four parameters, each of which has a 50-50 chance of being a certain value. The probability that an agent has any combination of four particular values is one in sixteen. When counting several thousand agents, the inferred probability should be approximately this value. However, suppose that there are the same four values, but this time the probability is split 95-5. In this case, the chance that the agent will have all four of the least common parameters is much much smaller than one part in several thousand. A counting strategy might turn up one or two agents, but would be likely to overestimate or underestimate the true probability. Thus, when dealing with small data sets (relative to the probability estimates), the enumeration strategy may not work effectively.

As alternative approach is to employ axioms of probabilities when performing the calculation. This approach is summarized in equation 7.

$$p(param_1 ... param_n) = p(p_1 ... p_n | literacy) * p(literacy) + p(p_1 ... p_n | \neg literacy) * p(\neg literacy) \quad (7)$$

Here, the probability for calculating $p(param_1 ... param_n)$ is done in two steps. The first half calculates the probability that the agent has attributes $p(param_1 ... param_n)$ and is literate; the second half calculates the probability of the attributes and that the agent is illiterate. Since there can only be two values for literacy, this approach effectively sums over all possible values of literacy and creates the desired result.

The first value on the left-hand side of the equation, $p(param_1 | literacy) ... p(param_n | literacy) * p(literacy)$, is the same value that is calculated in section 4.3. Thus, the calculation of this value requires no additional work. The other value, $p(param_1 | literacy) ... p(param_n | \neg literacy) * p(\neg literacy)$, requires a slight amount of additional effort; however, most of the work is already done.

Since literacy can only be true or false, $p(\neg literacy)$ is simply $1 - p(literacy)$. This calculation is relatively simple.

The other part of the formula, $p(param_1 ... param_n | \neg literacy)$, can be calculated in a fashion similar to that of section 4.3. First, the expression must be broken down using the naive Bayes assumption, as per equation 8.

$$p(param_1 ... param_n | \neg literacy) = \sum_{i=1}^{n} p(param_1 | \neg literacy) \quad (8)$$

As before, it is not possible to calculate $p(param_1 | \neg literacy)$. It is necessary to transform this equation by an application of Bayes Rule, resulting in equation 9.

$$p(param_i | \neg literacy) = \frac{\neg literacy | p(param_i) * p(param_i)}{p(\neg literacy_i)} \quad (9)$$

Table 1: Example input file

Gender	Male	85%
Gender	Female	88%
Age	[0-29]	89%
Age	[30-39]	88%
...	...	...

All the parameters are known here. The expression $p(\neg literacy|param_i)$ is the negated value that was supplied to Construct as $p(literacy|param_i)$. The value $p(param_i$ was already inferred for the calculation of other steps. Lastly, the value $p(\neg literacy)$ was already calculated and used several times before.

While it may seem as if this method of calculating the $p(param_1...param_n)$ is long and involved, it can actually be performed concurrently with the calculation of $p(param_1...param_n|literacy)$. Calculation using this method can actually much more rapid than calculation via enumeration.

4.4 Putting it together

Using the values calculated in the previous three sections, it is possible to calculate equation 2, the overall probability of $p(literacy|param_1...paramn)$. Probabilities will be between zero and one, though there may be some difficulties due to the averaging of the literacy probabilities in the creation of $p(literacy)$.

The probabilities generated will approximate the conditional input values supplied to Construct. However, there are two reasons for this difference. First, since the formula only creates a probability, it is possible for biases to occur based on different random numbers being selected, especially when there are only a small number of agents or few agents with a particular attribute. Second, due to the averaging necessary in the calculations of section 4.1, the values are unlikely to be exact unless the literacy probabilities are equal over all parameters. Most values, however, should be within five percentage points of the values specified in the input files.

The entire calculation of $p(literacy|param_1...param_n)$ takes less than a second to perform on a 2.19Ghz machine with 8GB of RAM should the five parameters used in table 2 be used in the calculation. In a typical virtual experiment lasting ten to twenty minutes, this additional computation represents less than a fraction of one percent of the total running time. Calculating any of the access probabilities requires analogous calculations and can be performed in approximately the same time.

5 Input and Use

In order to use the logic described in section 4, several types of input are necessary. These are described in the following sections

5.1 Input file

To enable the literacy or access logic, it is necessary to include the input parameter “XXX_probability_file”, where “XXX” is “agent_literacy” for literacy and “internet_access” for internet access. Omitting this parameter should cause the most current versions of Construct (versions built after January 2008) to crash, since it is a required parameter. However, it may be the case that the parameter is set to be non-essential, in which case the literacy modifications will be silently omitted.

The value of the parameter "agent_literacy_probability_file" or "internet_access_probability_file" should be the path to a .csv – comma separated value file – of literacy probabilities. This file is required at runtime. Either relative or absolute paths can be specified; if only a filename is given then Construct will look in the same folder as the .xml input file (which is not necessarily the folder with the executable). If the file cannot be found, then Construct will exit with an error message.

An example format of the comma-separated value file is given in figure 1. The file consists of three columns. The first column contains the name of the attribute that is being described, the second column contains the attribute value, and the third column the associated literacy probability. Probabilities must be specified between zero and one; values outside this range are unexpected and will cause the program to crash.

It is possible for the literacy file to perform inference using attribute ranges. For instance, it is possible to set the attribute value to be a range, as is seen for the age attribute in figure 1. The syntax for this is an open bracket [, followed by a number, followed by a hyphen, followed by another number, followed by a close bracket]. This will match any attribute within the range (inclusive). Such behavior significantly decreases the size of the file, especially when dealing with continuous or discrete numeric values.

There is no header row in the .csv file; data should be placed starting in row one. Blank lines will be skipped, and it is often convenient to skip lines in the .csv file to improve readability. While not required, literacy values corresponding to the same attributes should be kept together for easier interpretation and change. Though there may eventually be a commenting mechanism put in place in order to allow user comments in the .csv file, this feature is not supported at this time. Users are suggested to keep their comments in an associated text file or in the .xml input deck directly.

5.2 Agent parameters

To have an agent use the literacy mechanism when communicating, it is necessary to add the parameter "literacy_required" to an agent. When this parameter is set, all future interactions with this agent will perform literacy checks for the partner agent. The "literacy_required" parameter will most often be set for agents such as books, web pages, fliers, and other such agents which require some level of reading comprehension for interaction. Similar per-agent parameters such as "access_required" will force agents to have the access parameter set in order to be able to communicate with the agent.

It is most common for parameters such as "literacy_required" to be set in the "agent_types" area of the Construct input file. This will ensure that the parameter propagates to all agents which are instances of that agent type. However, it can be set on a per-agent basis to allow for additional user flexibility.

Note that parameters such as "literacy_required" define whether a not the literacy attenuation (or access requirement, or other factors) are in place for the agent's *interaction partner*, not for the agent itself. Interaction behavior will depend upon the interaction partner's value for the inferred literacy, access, or other value; for instance, an agent with the "literacy_required" parameter set will not modify its communication when dealing with a literate interaction partner.

A "use literacy" or other such parameter is not required for the interaction partner that will be potentially affected by literacy. If the literacy mechanism (or other access mechanism) is enabled for the simulation, the mechanism will automatically display its effect if an agent begins to interact with an agent that requires literacy or access. No specific parameter is required for the potential partner, though it will be necessary to either set or infer a literacy (or access) parameter for the agent.

Table 2: Prose literacy rates in America, 2003

Attribute	**Value**		**Percent above "below basic" literacy**
Gender	Male		85%
	Female		88%
Income	<$30k	*	82%
	$30k-$50k	*	86%
	$50k-$75k	*	92%
	>$75k	*	99%
Age	<30	*	89%
	30-40	*	88%
	40-50		89%
	50-65		87%
	65+		77%
Race	White		93%
	Black		76%
	Asian		86%
	Hispanic		56%
	Others	*	85%
Education	Less than hs	*	50%
	High school	*	86%
	College	*	93%
	Graduate		99%
Average			86%

* averaged or re-aggregated for Construct binning

5.3 Simulation parameters

While the above two sets of parameters completely describe what is necessary to enable the access mechanism, the literacy mechanism is more complex and relies on two additional simulation-level parameters. These parameters govern how much information is both correctly and incorrectly conveyed to illiterate agents if they are able (and choose to) interact with materials requiring literacy.

The "literacy_coverage_mean" parameter determines what percent of the information will not be absorbed by the agent at all. This parameter acts by shortening the message conveyed to the illiterate recipient; this parameter reflects the fact that illiterate agents are not able to absorb as much information from printed materials, but as long as this parameter is less than one some information will be absorbed (from pictures and other parts of the media). The "literacy_coverage_accuracy" parameter determines whether or not the agent was able to infer the correct information from the text. If an illiterate agent draws a random number less than this parameter, the Construct system will remove an additional portion of the message information and replace it with a random fact drawn from the total possible number of simulation facts. This process represents the global misinformation effect in which the agent believes it is reading one thing when it is actually reading something completely different.

6 Grounding Data

In order to make the newly-implemented mechanism as useful as possible, a literature search was conducted in order to obtain accurate information about American literacy and access rates. A variety of sources were examined, including data gathered by official government agencies, research foundations, and industry groups. The data gathered in this section present the results of this search. Some of the data had to be re-binned for use with the socio-demographic information commonly used in Construct; such data are marked in the corresponding tables. The sources for the data are available in the references of this document, and to the best of the authors' knowledge, is the most current data available as of this writing.

This section is divided into four sub-sections presenting the results of four different literature searches. Section 6.1 presents the results of the search corresponding to literacy rates in America. Section 6.2 presents data focusing on the consequences of illiteracy with respect to misinformation. Sections 6.3 and 6.4 present data that discusses the prevalence of Internet access and newspaper readership, respectively.

6.1 Literacy data

In the United States, literacy is not universal. Though the literacy question was removed from the 1940 census because it was assumed that "most people could read and write", this assumption was challenged twenty years later [12]. A literacy-related question was reinstated in the 1970 census, but rephrased as a years-of-education question; the literacy rate was determined to be 99% [12]. However, with comprehensive studies on literacy begun during the last quarter of the twentieth century, a more troubling picture revealed. Today's statistics present a much more nuanced view of reading skills by breaking it down into five levels, the lowest being "below basic" literacy. Individuals with below basic literacy may not be able to find information in short, commonplace prose texts, following written instructions in prose documents, signing and dating a document, or locating numbers in order to perform simple additive operations [13, 4]. The most comprehensive survey to date suggests that thirteen percent of the American population functions at this "below basic" level, suggesting that the national literacy rate in the United States 86% as of 2003. According to the most recent findings, then, one in eight Americans is functionally illiterate [4].

While the overall literacy rate for the nation is 86%, it varies considerably by demographic variable; table 2 reports the number of agents who scored above "below basic" literacy by socio-demographic characteristic. As the table suggests, men are more likely to be illiterate than women: only 85% of men are literate while 88% of women are [4]. Larger differences are seen by education level, income, age, race, and native language. Increased education is highly corralated with increased literacy. For instance, only 50% of those having only a primary school education are functionally literate according to the NAAL survey, as compared to 86% of those with a high school degree, 93% with some college or college degree, and 99+% for those with some post-graduate education [4]. Literacy is also positively correlated with income. About 82% of peple making less than $30,000 a year are literate, which rises to 86% of people making $30,000-50,000, to 92% for people making $50,000-75,000, and to 99+% for those making more than $75,000 [4]. On the other hand, literacy is negatively correlated with age. While the literacy rate is 89% for individuals for young people, the rate is 88% for the majority of the rest of the adult population – though only 77% for those older than 65 [4]. Race has also been shown to have an important effect on literacy rates: while whites have an average literacy rate of 93% for all ages and incomes, only 86% of asians are literate (in English), 76% of blacks, and 56% of hispanics (again in English), and 85% for those of other ethnicities [4]. However, much of the above data can be explained by looking at the literacy rates for native and non-native speakers of English. While the literacy rate for those brought up with English as their primary language is 91%, it is 86% for those brought up speaking both English and Spanish and 39% for those who grew up speaking only Spanish [4]. As the findings of the NAAL survey only examined proficiency in English, it is likely that individuals may be literate in another language; however, as Construct does not yet model language spoken, this parameter is not yet used.

These values from the 2003 NAAL survey, by socio-demographic, have been included in files external to the Construct binary but in common use in Construct experiments. There may be several other important sources of general literacy data, but the most comprehensive found – and the ones used in Construct experiments – use this NAAL study data. Some experimenters, however, may have more specific data if working with a particular region or sub-population; as described later, it is possible to use such data in place of the NAAL data described here.

6.2 Misinformation data

While it is convenient for the purposes of this simulation to hermetically divide the "literate" and "illiterate", this is not exactly the case. People who are considered literate can often make mistakes on their interpretation of a docu-

Table 3: Probabilities of Successful Completion of a Reading Task, 2001

Document sophistication	Below "below basic" literacy	Just above "below basic"
Most simple	81%	95%
Simple	40%	76%
Average	18%	46%
Complex	7%	21%
Most complex	6%	18%

* See Kirsch '01 for full table and details

ment, and illiterate individuals are often able to glean some information from a printed page. While the designers and implementers of Construct have made a conscious decision to ignore the former effect at this time, as discussed in section 2.3, the second has been implemented as a misinformation mechanism.

Table 3, adapted from Kirsch '01, was gathered from various literacy experiments taken for the international adult literacy and life skills survey (AILS) [14]. It presents the percentage of responses that individuals with "below basic" literacy were able to correctly identify on a multiple-choice exam. As discussed in section 6.1, the proportion of illiterate agents was chosen to match the percentage of the population who had below basic literacy. These individuals often had a substantial amount of difficulty with simple tasks, such as finding the total on a bill. Table 3 suggests that agents at the below-basic borderline have between a 81% and a 95% chance of giving the correct answer to a simple problem, and between a 40% and 76% chance for a simple one. Since the percentage that were not correct were assumed to be mistakes, this meant that approximately 10% of the excersizes involving the most simple level of documents were not answered correctly, and that nearly fifty percent were not answered correctly for documents with the simple level of difficulty. More difficult documents were rarely answered correctly.

In most Construct experiments, a value of 50% is used at the misinformation rate; this number is roughly mid-way between the numbers given for the "simple" document. This is done under the assumption that the printed materials are designed for a relatively diverse audience. If the printed materials are assumed to be more complex than usual, this number can be lowered in accordance with table 3; values as low as 30% would not be surprising if more sophisticated educational interventions were to be used.

It is recognized that the data presented in table 3 is not perfect – it was taken from a controlled and potentially artificial environment, that it is not domain specific, and other such criticisms are certainly valid. Nevertheless, it represents the best publicly-available data for misinformation rates. Since these numbers are a simulation parameter, it can be changed if other experimenters are able to locate domain-specific data or a more up-to-date information source.

6.3 Internet access data

In the last quarter century, the rapid rise of the Internet has provided a host of new options for individuals seeking information [6]. The ability to view pages on the web, as well as to exchange email or search, have made online access increasingly important in modern society. While the vast majority of Americans have been able to take advantage of online material, and many do so on a regular basis, a significant percentage of the population continues to lack regular access to the Internet [6]. While there are individuals of every socio-economic strata that do not take advantage of online content, recent studies have shown that a number of population groups do not use the web on a regular basis – or lack Internet access altogether.

Table 4 presents a subset of the material pertaining to the socio-demographic characteristics used in Construct simulation. The information presented in the table was gathered from the Pew Internet and American Life Project, a non-profit group dedicated to studying the Internet and online behavior [6]. The Pew group reported that 70% of Americans use the Internet "at least occasionally" as of the middle of 2007, though this number has been rising rapidly from

Table 4: Internet access statistics in America, 2007

Attribute	Value		Percent who use the Internet "at least occasionally"
Gender	Male		71%
	Female		70%
Income	<$30k	*	55%
	$30k-$50k	*	69%
	$50k-$75k	*	82%
	>$75k	*	93%
Age	<30	*	90%
	30-40	*	83%
	40-50		83%
	50-65		65%
	65+		32%
Race	White		75%
	Black		68%
	Asian		**%
	Hispanic		74%
	Others		**%
Education	Less than hs	*	40%
	High school	*	61%
	College	*	81%
	Graduate		91%
Average			70%

* averaged or re-aggregated for Construct binning
** no data available; average US value of 70% used

year to year and was projected to cross 75% in early 2008 [6]. Men were slightly more likely to have used the Internet, but the difference was slight and well within the poll's margin of error [15]. Strong differences were found by income bracket, with only about half of those making less than $30,000 a year using the Internet occasionally while more than nine in ten of those making at least $75,000 did so [15]. Perhaps predictably, Internet access was most common among the young: while about 90% of young people used the Internet, less than one in three seniors did so [15]. Whites were more likely to use the Internet than those of other races; while the Pew trust had focus groups studying the use of the Internet by Black and Hispanic groups, they did not collect data for individuals of other ethnicities [16]. Lastly, income was strongly correlated with Internet use, as barely 40% of those with less than a high school diploma use the Internet occasionally as compared to 91% of those with masters, doctorates, or other post-baccalaureate schooling [15].

These values from the Pew surveys have been included in files external to the Construct binary but in common use in Construct experiments. As with the literacy data presented in section 6.1, the values in the external file are not fixed and can be changed if the experimenter chooses. Given the rapidly changing demography of Internet users, it is likely that this file will have to change rapidly in order to accurately reflect contemporary society. However, should an experimenter wish to change the percentages in order to examine arbitrary socio-demographic data, or because data is available for a specific sub-population, the external file facilitates such changes and allows for experiments to be run in a rapid and straightforward manner.

6.4 Readership data

Newspapers and printed pamphlets have been among the oldest forms of mass media, with histories that go back hundreds of years. By the time of the American revolution, there were well over forty newspapers available in what would

Table 5: Newspaper subscription statistics in America, 2007

Attribute	**Value**		**Percent who are reached by "one issue" of a newspaper**
Gender	Male		49%
	Female		44%
Income	<$30k	*	40%
	$30k-$50k	*	45%
	$50k-$75k	*	47%
	>$75k		58%
Age	<30	*	32%
	30-40		41%
	40-50		41%
	50-65		53%
	65+		58%
Race	White		49%
	Black		43%
	Asian		38%
	Hispanic		27%
	Others	*	31%
Education	Less than hs	*	29%
	High school	*	44%
	College	*	57%
	Graduate		57%
Average			46%

* averaged or re-aggregated for Construct binning

become the United States; since that time, that number has grown substantially. However, in the last half-century that number has begun to decrease, driven by the easy access to new media types such as television, radio, and (most recently) the Internet [17]. The absolute number of newspaper subscribers has not changed substantially since the 1950s, and there has been a marked decrease in the number of daily newspapers published in the United States during that time as a relatively smaller percentage of the country reads a daily paper [18]. In addition to the fact individuals of different regions tend to obtain their news via different papers – unlike the Internet, newspapers are tied much closer to geography – there are substantial variations via socio-demographics.

Table 5 presents statistics on newspaper readership in the United States as of 2007 [17]. The statistics presented are the percentages of Americas who read a single issue of a daily newspaper; while statistics are available for the number of individuals who read one issue of a weekday newspaper during a workweek or the number of individuals who read a Sunday paper, the single-day statistic was decided upon to single out those who might be more likely to read a newspaper carefully enough to notice a print advertisement in the media. As can be seen in the table, there are noticeable socio-demographic changes with regard to readership. Men are much more likely to read a daily newspaper than are women; the gender difference is nearly five percentage points. High income individuals are more likely to read newspapers than are low-income individuals; the jump is particularly marked for those making more than $75,000 a year, 58% of whom read a daily. Older Americans are more likely to read the newsaper, with over six in ten seniors reading a paper but barely four in ten of those under 30. Whites are also more likely to read newspapers than Blacks, Asians or Hispanics; while nearly half of all Whites read a daily, barely a quarter of Hispanics do so. Increased schooling is also strongly associated with daily newspaper readership, with 29% of those with less than a high school education reading a daily newspaper as compared to 57% of those with post-baccalaureate education doing so. While it is recognized that some of these readership trends are strongly correlated with literacy, there are substantial differences between literacy and newspaper readership since not all literate individuals read daily newspapers and some functionally illiterate people are able to obtain some amount of information from printed news sources.

Research has indicated that individuals of different socio-demographic groups are more likely to read certain kinds of newspapers; for instance, men are more likely to read newsprints related to sports and high-income people are more likely to read financial sections [19]. Similarly, research has suggested that individuals may read different sections of the paper, and therefore receive different kinds of information from the same newspaper. However, for simplicity in modeling and because the effect was not thought to be critical at this stage of model design, this information was duly noted but not implemented in the Construct simulation [19]. Future versions of the tool may modify this feature to more accurately reflect the reading habits of different socio-demographic groups.

These values from the NAA surveys have been included in files external to the Construct binary but in common use in Construct experiments. As with the other types of data presented in this section, the values in the external file are not fixed and can be changed if the experimenter chooses. Again, should an experimenter wish to change the percentages in order to examine arbitrary socio-demographic data, or because data is available for a specific sub-population, the external file facilitates such changes and allows for experiments to be run in a rapid and straightforward manner.

7 Example Experiment

To facilitate the design of future experiments using these implemented literacy and access mechanisms, and to more fully describe the Construct experimental procedure, it was decided to investigate the effect of literacy on the effect of multiple interventions. The following subsections describe the experimental setup, discuss the results of the performed experiment, and briefly summarize the results as pertaining to the effects of literacy and information access.

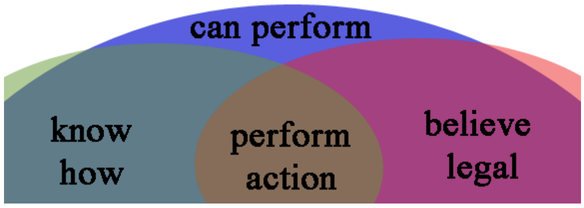
7.1 Experiment description

The experiment and data presented here focus on the number of agents in the society who believe a certain behavior is acceptable. Such modeling has served as a sub-portion of previously-modeled virtual experiments, such as experiments dealing with the propagation of information relating to taxpaying behavior in society [5, 20]. While other work has looked at the overall effect of interventions or of societal organization, the experiment here examines the effect of literacy and information access parameters on past work. Specifically, this virtual experiment examines the effect of literacy and access parameters on two different educational interventions, a print advertisement and a web page. The experimental design matrix is presented in table 6.

In the virtual experiment, a sample society of three thousand simulated agents was infiltrated by a promoting agent. This promoter sought to influence taxpaying behavior; this behavior can be thought of as convincing individuals to participate in an illegal tax scheme, to fail to take a legitimate tax credit, or to file tax forms in an incorrect fashion. The agents in the standard society interacted via homophily, as in past Construct experiments; however, their conversations were over-sampled to ensure that conversations of interest – namely, conversations in which taxpaying behavior information was conveyed – were highlighted over the course of the simulation. The main experiment, which was imagined to last a simulated year with semi-weekly interactions (104 total time periods), consisted of the promoter attempting to convince as many agents as possible to alter taxpaying behavior by providing knowledge and belief information to agents in the society. The action modeled was designed primarily for low-income individuals in the society since such individuals were more likely to be affected by the literacy and access features newly implemented in Construct.

Of the three thousand agents in the society, about 2200 of them were low-income agents who could potentially have their behavior influenced. These were the agents in the blue area of figure 3, the agents that could potentially perform the action. Agents could then learn information about the action via direct communication with the promoter, direct communication with an intervention, or via indirect communication via a chain of agents who talked to either the promoter or intervention. Agents could also begin to believe that the behavior was legal by beginning with an intrinsic bias, becoming convinced after communication the promoter, learning incorrect information from an intervention, deferring to influential agents in their social network if they had no other information, or by communicating with other agents who had one or more of these actions. Agents who actually took the action were low-income agents who knew sufficient information about the action and believed that the promoter's action was legal. Agents who lacked

Figure 3: The qualities of agents who perform the behavior



one or more of these criteria were still of interest, but were not necessarily ones whose behavior would be modified.

The setup of figure 3 suggested that there was a hierarchy of the agents of interest. For instance, all agents who changed their behavior believed that the behavior was legal, but only some of those who believed the behavior was legal actually acted on that belief. If an agent believed that the action was legal but did not take the action, it might not have known enough about the behavior to actually take action or might not have been an eligible low-income agent.

In this experiment, a decision was made to over-sample for low-income agents in order to increase the possible number of agents who took the action. This was done for statistical reasons as well as to have a sufficient number of matching in order to draw conclusions about the behavior of socio-demographic groups. All percentage statistics presented about this experiment are presented in terms of the number of low-income agents who were eligible to take the action, not of the total number of agents in the society. This prevents the statistics from being deflated by agents who could not have taken the action anyway and were not of interest to the promoter.

While the agents in the society initially started out with a largely negative attitude toward believing the behavior was legal, the promoter was occasionally able to change the beliefs of a substantial number of agents. For this reason, two educational interventions were examined:

- An advertisement, which sought to provide a small amount of knowledge and belief information but to distribute this information to a wide audience. While it did not offer much in the way of knowledge or belief, it was available in newspapers and potentially readable by a large number of agents – especially those who might not have any direct interest in performing the action but might find themselves contacted by the promoter in the future.
- A web page, which sought to provide a large amount of knowledge and belief information to a targeted audience. Though it could provide a large amount of belief information stating that the behavior was illegal, it could also provide a substantial amount of knowledge. Thus, it was possible for some agents to seek out the web site intervention and misconstrue the web site's message: some agents might learn how to perform the action from the web site and ignore any information regarding illegality. Additionally, the web site was primarily geared toward those who already knew something about the behavior so as to be a more focused way of deterring at-risk individuals from performing the action.

As described thus far, the experiment was similar to those types of experiments run in the past and is presented in the column labeled "no change" in table 6. However, this experiment differs from previous work because of the new Construct features modeled: the advertisement was affected by literacy as well as newspaper readership, while the web page was affected by literacy as well as Internet access. The results of the virtual experiment described here sought to tease out the effects of each of these parameters on intervention strength: modeling each of these effects more accurately would affect the strength of the intervention and lead to different behaviors in the society. The three mechanisms were tested individually – the middle columns in table 6 – and then were examined collectively. This last column was the most realistic experimental setting examined, in which any applicable subset of literacy, Internet access, and newspaper readership were used to more accurately measure the effects of the interventions applied to the society.

Table 6: Experiment description

Case	no changes	w/ literacy	w/ internet	w/ readership	most realistic case
No intervention	250 reps	N/A	N/A	N/A	N/A
Advertisement	250 reps	250 reps	N/A	250 reps	250 reps *
Web site	250 reps	250 reps	250 reps	N/A	250 reps †
Both ad and web	250 reps	250 reps	250 reps	250 reps	250 reps ‡

* literacy, readership enabled † literacy, internet access enabled ‡ all three enabled

The virtual experiment run was a modified version of a 2 x 2 x 5 cross, with two parameters for the advertisement (*not present* or *present*), two parameters for the web site *not present* or *present*), and five for the different access mechanisms (*none present*, *literacy only*, *Internet access only*, *newspaper readership only*, and *all three mechanisms*). However, since some of these experiments were not relevant (for instance, it was not necessary to examine the effect of modeling Internet access if no web site was present), only fourteen of the possible twenty cases were executed. This experimental design is summarized in table 6. Two hundred and fifty iterations were run for each relevant experimental case, resulting in 3,500 experimental runs. Since this analysis was performed as part of a larger package of experiments, the exact running time is not known; post-hoc, running time was estimated to be four days on an 4 x 2.4Ghz server with 64GB of memory. A subsequent set of runs was performed on the Purdue TeraGrid cluster system, a heterogeneous set of machines made available through the National Science Foundation [21]. This follow-up set of runs took about twelve hours to complete. A more detailed description of the experiment parameters can be found in the description to virtual experiment Ia in the technical report "Information Access Experiments using Construct" [20].

7.2 Experimental results

Table 7 presents the results from the most realistic case modeled. The most realistic cases were the advertisement with literacy and readership, the web site with both literacy and Internet access, and the combination case with literacy, Internet access, and newspaper readership modeled. The table presents the percentage of eligible agents in the society who had the requisite knowledge, had requisite belief, and who decided to take the action. Since some agents were not eligible to take the action because their income was too high, the agents were ignored from the subsequent analysis and were not analyzed. The values in the table were calculated using the following helper formulas:

- The intermediate value MIR_{ki} was the mean percent who did k (either knew enough information to act on the behavior, believed the behavior was legal, or actually took action) when intervention i was present and the more realistic model of agent ability and access to information is used, i.e., only cognitive and information access constraints of relevance were turned on.
- The intermediate value MIU_{ki} was the mean percent who did k (either knew enough information to act on the behavior, believed the behavior was legal, or actually took action) when intervention i was present and the least realistic model of agent ability and access to information is used, i.e., there were no constraints on literacy or access.
- The intermediate value $M0_k$ was the mean percent who did k (either knew enough information to act on the behavior, believed the behavior was legal, or actually took action) when no intervention was present. It is important to note that when there was no intervention agent ability and access constraints that were relevant in this case irrelevant; i.e., the realistic and unrealistic case were identical.

Using these values, three derived values were created: the mean rate of activity (mean), the percentage increase in the activity (% inc), and the percentage change in activity (% chg). These values were defined as follows.

- The mean (mean) indicated the percentage of agents which performed the behavior. This was the value MIR_{ki} for the most realistic models of the interventions only interventions only.

- The percentage increase (% inc) indicated how much more – or less – behavior occurred when that intervention was and the more realistic model of agent ability and access to information is used as compared to a baseline case in which there is no intervention modeled. This was calculated as:

$$100 * (MIR_{ki} - M0_k)/M0_k \tag{10}$$

- The percentage change (% chg) was percentage change, for the impact of that intervention, when moving from the least realistic model of agent ability and access to information and the more realistic model of agent ability and access to information. This was calculated as:

$$100 * (MIR_{ki} - MIU_{ki})/MIU_{ki} \tag{11}$$

Several trends are apparent from the data in table 7. First, knowledge was highest when the web site was active as compared to when either the advertisement alone or the advertisement and web page were both active. When the web site alone was active, a mean of 3.21% of eligible agents knew how to do perform the action in the most realistic case, while with the other interventions a mean of only 3.15% knew enough information to take the action in the most realistic cases. Thus, two of the three cases had lower knowledge levels than that observed in the no intervention case – both the advertisement and the advertisement with web page had a 2% decrease. More importantly, though, there were important changes between the realistic cases modeled here and the simple cases as modeled in previous Construct runs. The most realistic advertisement case lead to slightly fewer agents knowing how to do the action as compared to the least realistic case, a decrease of 0.7%. On the other hand, the most realistic web site lead to an increase in the number of agents who knew how to do the action as compared to the least realistic case, wher 2.1% more agents knew how to perform the behavior. When both the ad and the web were active, 1.5% fewer agents knew know to do the action in the most realistic case as compared to the least realistic case. This meant that the literacy and readership mechanism lead to a further decrease in knowledge for the advertisement case when compared to the no literacy or access case, the literacy and internet access mechanisms actually contributed to an increase in knowledge for the web site. The combined case was similar to the web site in the sense that the most realistic case lead to a decrease in knowledge as compared to the least realistic case; even though the raw percentage of knowledgeable agents was roughly similar to the case with the advertisement alone, the base knowledge rate was much higher in the least realistic combined intervention case and thus lead to a larger percentage drop. These trends, and several others, will be discussed in more detail in subsequent sections.

While the effects of the literacy and access mechanisms on knowledge were mixed, the effects on belief were more consistent. Table 7 suggests that the number of matching agents who believed the action was legal was 3.53% in the most realistic advertisement case, a number that fell to 3.30% for the most realistic model of the web site and then to 3.23% when both the advertisement and the web site were active. While the advertisement lead to a 1.6% decrease in the percentage of agents beliving that the behavior was legal as compared to the no-intervention case, this became a 7.8% decrease with the web site and a nearly 10% decrease with the combined advertisement and web site. Taken together, this suggested that all three interventions were effective in reducing the number of agents who believed that the action was legal even when the most realistic mechanisms were used. However, these literacy and access mechanisms served to decrease the effectiveness of the interventions. The most realistic model of the advertisement – the advertisement modeling both literacy and newspaper access – lead to 0.7% more agents believing that the action was legal as compared to the simple intervention that did not, a factor that suggested that the additional cognitive and access mechanisms weakened the effect of the advertisement on belief. The effects for the web site, a 6.8% increase in behavior, and for the advertisement and web site, a 4.5% increase in the number of agents who performed the action, suggeseted that the more realistic models of these interventions also decreased the intervention's effectiveness at deterring belief.

An agent decided to perform the behavior when it both knew enough information about the behavior in order to act and believed the behavior was legal. In the most realistic advertisement case, nearly 0.6% of potentially qualifying agents went on to perform the action, while in the most realistic web site and most realistic combined case only 0.55% did so. While the 0.59% of the population that took the action for the advertisement was greater than the equivalent rate without any advertisement, it also marked a 7.1% increase over the least realitic advertisement rate. This suggested that the literacy and access mechanisms lead to an increase in the number of agents who took the action, and

Table 7: Comparison of most realistic cases to least realistic cases

Case	**know how**			**believe legal**			**perform**		
	mean	**% inc**	**% chg**	**mean**	**% inc**	**% chg**	**mean**	**% inc**	**% chg**
Advertisement	3.15%	-1.9%	-0.7%	3.53%	-1.6%	+0.7%	0.59%	+1.5%	+7.1%
Web site	3.21%	-0.1%	+2.1%	3.30%	-7.8%	+6.8%	0.55%	-5.2%	+1.3%
Both ad and web	3.15%	-1.9%	-1.5%	3.23%	-9.9%	+4.5%	0.55%	-5.1%	+0.1%

this increase was substantial enough to lead the advertisement to increase the rate of overall rate of behavior in the society. The web site lead to a 5.2% fall in behavior relative to the no intervention case even though the most realistic model of the web site lead to 1.3% fewer agents performed the action than in the least realistic case. The web site, then, caused a decrease in behavior even though this effect was weakened by the cognitive and access mechanisms. Lastly, while the most realistic model of the combined advertisement and web site case lead to a 5.1% decrease in behavior, it had close to no change when compared with the least realistic case. While the combined case did lead to a decrease in behavior, the effect of the cognitive mechanisms on this case was as strong as it was for either of the interventions individually.

In order to place the results of table 7 in context, it is necessary to examine the effects of the individual literacy and access mechanisms on the interventions. This additional data is presented in tables 8, 9, and 10. In these tables, the results are presented for each of the cognitive and access effects separately before being used in the most realistic case. These tables contain the following values:

- The mean number of agents taking the action (mean) is again displayed as a percentage of the number of low-income agents who have the particular quality in a matter analogous to table 7. This is again the MIR_{ki} value, but this time it could take on the value for any relevant mechanism on intervention i.
- The standard deviation (std dev) about the mean MIR_{ki} is presented as a range about this mean in order to convey information about the variability in the simulation results. Since many aspects of the simulation can lead to variability – from changes in who could speak to an intervention to what sort of information is spread from agent to agent – there could be substantial differences between runs; the standard deviation captures this information as a measure of the volatility of the intervention or realism scenario to various simulation conditions.
- Lastly, the percentage change (% chg) was again presented; it is the percentage change observed when moving from the least realistic model of the intervention to the most realistic model. This was calculated as per equation 11, $100 * (MIR_{ki} - MIU_{ki})/MIU_{ki}$. In this equation, MIR_{ki} could take on the value for any relevant mechanism on intervention i and MIU_{ki} would take on the value of the least realistic intervention modeled – the intervention without any cognitive or access mechanisms active.

Sections 7.3, 7.4, and 7.5 discuss these tables in turn.

7.3 Advertisement experiment results

When compared to the no-intervention case, the least realistic model of the advertisement lead to a slight decrease in the amount of knowledge available in the society, a very slight decrease in belief, as well as a drop in the number who performed the action. While this drop in the number of agents who modified their behavior was not necessarily seen when the same model was run as in earlier work – previous models had suggested no change in the number who changed their behavior when the advertisement was present – the drop observed here was the smallest of the drops observed of any of the interventions.

In comparing the effects of the different cognitive and access mechanisms to this least realistic model of the advertisement, as shown in table 8, several trends were apparent. When literacy was modeled, there was a slight decrease in the number of agents who knew how to do the action, though there was a much larger decrease in the number of

Table 8: Advertisement results

Advertisement	know how			believe legal			perform		
	mean	**std dev**	**% chg**	**mean**	**std dev**	**% chg**	**mean**	**std dev**	**% chg**
No mechanisms	3.17%	0.54%	—	3.50%	0.61%	—	0.55%	0.16%	—
w/ literacy	3.16%	0.56%	-0.4%	3.46%	0.54%	-1.1%	0.58%	0.16%	+6.2%
w/ readership	3.22%	0.68%	+1.2%	3.54%	0.60%	+1.1%	0.58%	0.19%	+5.1%
w/ both	3.15%	0.56%	-0.7%	3.53%	0.51%	+0.7%	0.59%	0.16%	+7.1%

Table 9: Web page results

Web site	know how			believe legal			performed		
	mean	**std dev**	**% chg**	**mean**	**std dev**	**% chg**	**mean**	**std dev**	**% chg**
No mechanisms	3.15%	0.56%	—	3.09%	0.50%	—	0.54%	0.14%	—
w/ literacy	3.13%	0.55%	-0.6%	3.24%	0.51%	+4.9%	0.54%	0.17%	-1.5%
w/ access	3.06%	0.58%	-2.7%	3.25%	0.56%	+5.1%	0.55%	0.16%	+1.2%
w/ both	3.21%	0.52%	+2.1%	3.30%	0.52%	+6.8%	0.55%	0.16%	+1.3%

Table 10: Combined ad+web results

Both ad and web	know how			believe legal			perform		
	mean	**std dev**	**% chg**	**mean**	**std dev**	**% chg**	**mean**	**std dev**	**% chg**
No mechanisms	3.20%	0.50%	—	3.09%	0.43%	—	0.55%	0.18%	—
w/ literacy	3.21%	0.58%	+0.2%	3.15%	0.50%	+2.0%	0.54%	0.18%	-1.5%
w/ access	3.16%	0.53%	-1.3%	3.23%	0.51%	+4.7%	0.55%	0.17%	-0.6%
w/ readership	3.13%	0.54%	-2.1%	3.11%	0.50%	+0.6%	0.52%	0.18%	-4.4%
w/ all active	3.15%	0.59%	-1.5%	3.23%	0.61%	+4.5%	0.55%	0.18%	+0.1%

agents who knew enough information to take the action. This effect was largely due to the truncation and misinformation effects of illiteracy, factors which decreased the amount of behavior-related information sent to agents in the simulation. Literacy also contributed to a further decrease in the number of agents who believed that the action was legal, contributing to a 1.1% decrease relative to the least realistic case. While the decreases in both knowledge and belief were due partially to the direct effects of misinformation and truncation due to illiteracy, several other factors were at work as well: when literacy was active, there was less contact with the promoter due to the smaller number of agents who knew how to take the action, and there was less contact with the advertisement since fewer agents learned about the behavior and therefore were likely to pay attention to it.

Despite the aggregate decreases in knowledge and belief however, there was a net increase in the number who changed their behavior when literacy was active as compared to the baseline case without cognitive or access mechanisms modeled. Since agents who performed the action had to both know of the behavior and believe the behavior was legal, decreasing an individual effect such as knowledge or belief might or might not decrease the number who perform the action. By decreasing the number of agents who know of the behavior, for instance, it is possible – but not be guaranteed – to decrease the number of agents who perform the action. Thus, when decreasing the dark green area of agents who know how to do the action in figure 3, it is possible to decrease the green area without decreasing the size of the brown area. When literacy was modeled for the advertisement, this exact scenario occurred: belief and knowledge both decreased relative to the advertisement without literacy, but the number who performed the action increased. In fact, while the decreases in knowledge and belief were not statistically significant, the observed 6.2% inrease in the number who take action was statistically significant at the .01 level. Indeed, the increase may in some ways be related to the fact that the promoter communicated with fewer different agents, as it could spend additional effort contacting old agents and ensuring that they followed through on what they learned in an initial interaction.

With newspaper readership, there were different effects. When newspaper readership was implemented, the mean number of agents who accessed the advertisement dropped from nearly 8.5% of the population to 4.0%, a 56% drop. This lead to a decrease in the effectiveness of the advertisement: while in the no-readership case 3.17% of the society knew of the behavior, this number rose 1.2% to be 3.22% of the society. Belief also rose from 3.50% to 3.54%, suggesting that the readership mechanism lead to a decrease in the advertisement in terms of its effect on belief.

Most notably, though, the readership mechanism lead to a large increase in the observed standard deviations. The standard deviation in knowledge increased twenty-five percent to 0.68%, while the standard deviation for belief rose to nearly equal that of the no cognitive mechanism case (and substantially higher than the literacy case). This suggested that the effects of the readership modification were felt unevenly in the simulation: in the replications where readership had a greater effect, the effect of the advertisement was less pronounced and were similar to the no-intervention case, but in the replications where readership had a lesser effect, the results were more similar to the baseline case where the advertisement was modeled without readership. Additionally, though, this rise suggested that the intervention was less reliable overall and was therefore much more similar to the no-intervention case. The no intervention case had the highest standard deviation of all the cases modeled since that case had no outside forces (other than the promoter) to influence knowledge, belief, and behavior. In the no-intervention case, there was nothing to counteract the spread of knowledge or belief; this could lead to inconsistent spreading of information between different runs. The readership mechanism on the advertisement, which decreased the number of individuals who interacted with the advertisement, made the resulting standard deviations much more similar to the no-intervention in terms to the larger number of outliers.

The number who performed the action – those that had sufficient knowledge and belief – increased 5.1% when the readership mechanism was included. This effect was likely due to the direct effect of the increase in knowledge and belief in the society, leading to a greater number of agents who had both attributes. However, the standard deviation on the number who performed the action was also quite large, suggesting that there were large swings based on the size of the newspaper-reading community. In runs where greater numbers of agents knew of the behavior or believed that the behavior was legal, more agents took action, while increases in those metrics could explain about fifty percent of the variance observed in the number who took action, a significant amount of the variation was due to the individual traits and communications of the agents themselves.

When both literacy and readership were modeled in the same simulation run – the most realistic advertisement case modeled – different results were observed. Knowledge of how to perform the action fell further to 3.15%, a 0.7% decrease from the least realistic case. As in the literacy case, this effect was largely the effect of the truncation effect as well as the decreased number of agents who interacted with the promoter to learn information about the behavior relative to the least realistic case. Belief in the behavior increased slightly to 3.53%, suggesting that the literacy effect and the lack of access to the advertisement helped to create a situation where the promoter's message was more accepted than in the less realistic case.

Despite the decrease in knowledge, however, there was a noticeable increase in the number of agents who performed the action. The number of agents who chose to perform the action rose to 0.59%, the highest observed in any of the virtual experiments run in this simulation. This effect was 7.1% higher than the least realistic case and 1.5% higher than the no-intervention case, suggesting that the more realistic intervention had the ability to greatly increase the number of agents who acted on the promoter's information. In the most realistic case, the number who performed the action rose further than in any of the individual cases.

7.4 Web site experiment results

In contrast to the advertisement, which provided only a small amount of knowledge and would often have only a small effect on belief, the web site could provide a large amount of information while also providing a substantial amount of belief-related information to deter agents from performing the action pushed by the promoter. As compared to the no-intervention case, the web site without cognitive or access limitations contributed to decrease in knowledge as well as a substantial decrease in belief. This differed slightly from past work – which had suggested a slight increase in knowledge coupled with a large decrease in belief – which suggested that the web site in this case provided additional belief information at the expense of knowledge. Nevertheless, the number of agents who performed the action was also substantially less than that observed for the no intervention case, a finding that was proportional with the decrease observed in earlier data.

As presented in table 9, the more realistic literacy mechanism lead to slight changes from the least realistic model of the web site. When the literacy mechanism was enabled for the web site, an 0.6% decrease in the number who changed their behavior was observed when compared to the experiments without literacy. Thise effect was partially due to web site truncation misinformation, since agents communicating with the web site were less likely to learn how to perform the action from the intervention alone. On the other hand, the mean number of agents who believed the behavior was legal rose from 3.09% to 3.24%, an increase of nearly 5%. This, too, was a function of the truncation and misinformation effect: since the web site sent a shorter and less complete message, fewer agents learned sufficient belief information from the web site to change their views. While a substantial number of agents communicating with the web site were literate agents, and therefore did receive the full message, about forty percent of the agent population was illiterate and could suffer from the misinformation effect. It was primarily these forty percent that lead to a lower percentage of agents knowing how to do the behavior as well as believing that the behavior was legal, though the second-order effects of agents passing on the web site information to other agents was also diminished by the literacy restrictions.

As compared to the model without literacy, the fraction of agents who changed their behavior was 1.5% lower for the model with literacy. This was primarily due to the decrease in knowledge but also due to the fact that agents were affetted by the misinformation in different ways. The large increase in belief did not necessarily lead to a large increase in behavior for several reasons. The agents who believed the action was legel were not necessarily knowledgeable about how to perform the action and thus lacked a critical component necessary for changing their behavior. Additionally, though, some of who believed the behavior was legal obtained this information by communicating with the web site and obtained the wrong message due to the misinformation effect. When this occurred, it was unlikely for those agents to have learned sufficient information from the web site due to the truncation effects; these agents would obtain only belief information without obtaining sufficient knowledge about the behavior in order to take action.

When the Internet access mechanism was enabled, a 2.7% decrease in knowledge was observed relative to the no intervention case. Under the same conditions, though, a 5.1% increase in belief was observed. During this time, communication with the web site decreased markedly: the number of agents who communicated with the web site fell from 5.2% to 3.6%, a 45% decrease due to the more realistic model of Internet access. During that time, communication with the promoter decreased, a fact which – in conjunction with the decreased information being spread by the web site – explains the observed drop in knowledge. The increase in belief is due to the decreased communication with the web site, as the web site had less of an opportunity to change the belief in the society. Even though the promoter was less active and less likely to sway agents, the web site was substantially more affected by the more realistic access mechanism.

Also importantly, the standard deviations around these statistics was substantially higher than those observed for the other web site modifications. This increased variability was partially due to the extra randomness included due to minor changes in initial Internet access assignment, as an agent might be deterred by the web site in one run but lack Internet access in another. Like the newspaper mechanism for the advertisement, though, the Internet access mechanism restricted the number of agents who could interact with the mechanism. The observed standard deviations – 0.58% for knowledge and 0.56% for belief – was higher than that for the no mechanism case and much more similar to the no intervention case. The higher variance observed is consistent with the fact that fewer agents learned information from the intervention, leading to an increased number of replications in which there were high and low outliers.

When the Internet access mechanism was active, the fraction who took action increased by 1.2%. This is logical, as the lack of Internet access decreased the effectiveness of the intervention since fewer agents had access to the belief information there. The observed rise is largely due to an increase in belief in the society, although the rise is somewhat tempered by the decrease in knowledge caused by the lack of communication with the intervention and decreased promoter communication.

When both literacy and Internet access were modeled in the same simulation, the number of agents who knew how to do the action rose 2.1% to 3.21% of the eligible society. Part of the observed increase was due to an increase in interaction with the promoter; the mean number of promoter interaction was much higher in this scenario than in the least realistic scenario. Another part of this effect, however, may have been due to an increase in the amount of knowledge information relative to belief information spread by the web site; had the access mechanisms not been in place, it was possible that this swing might have caused a greater increase in knowledge. The increase may also have been associated with the corresponding increase in belief. This in turn was due to the modeled effects, the lack of access to or misrepresentation of the intervention. This factor may have drawn additional agents to speak with the promoter and learn facts related to the action. Belief rose in this case to 3.30%, an increase over the literacy case alone and the access case alone. While some of the rise in belief can be attributed to the increased interaction with the promoter, some was due to less anti-activity belief information being passed to the society, both as a function of the messages being sent by the web site and the information being lost to the lack of access or literacy of the receivers.

In the most realistic case modeled, the percentage of agents who took the action rate was 0.55%. This value was almost identical to the value observed for the access mechanism alone. This rise was primarily driven by both the increase in the number of agents who knew how to perform the action and the increase in the number of agents who believed the action was legal. Even though knowledge and belief both rose relative to the case in which only Internet access was modeled, the observed increase in the number who performed the action was very slight. This suggested that the literacy mechanism, and the increased interaction with the promoter, may have informed additional agents about the behavior but not lead to a large increase in the number of agents who both knew enough information and held the legality belief.

7.5 Combined experiment results

When the advertisement and web page were together, as presented in table 10, more information could be spread to the society, but additional belief information could also be disseminated in order to counteract the increased knowledge available. When the least realistic model of the advertisement and web site were both active in the same society, knowledge was increased as compared to the web page alone or the advertisement alone. Belief was much less than in the least realistic advertisement case but virtually unchanged from the web page. The number who changed their behavior was virtually unchanged relative to these two cases, though substantially less than the no intervention case. These results also differed slightly from earlier work. In previous work, both the number who performed the action and the number who knew how were slightly lower in the combined case than the case with the web page alone.

The literacy mechanism affected both the advertisement and the web site interventions. When the literacy mechanism was included, knowledge was virtually unchanged at 3.21% suggesting that the truncation and misinformation effects had almost no effect on the combined case. This was a different result than that observed for the web site alone or intervention alone, though the magnitude of the change was very small. Belief, however, was up slightly, suggesting that the literacy modification from the web page was the dominant factor. Since the web page provided the greatest amount of belief information, it was hit hardest by the change due to literacy. The behavioral effect of literacy on the advertisement was more than drowned out by the effect of the web page, as promoter contact (one of the reasons for the drop in belief in the advertisement case) was not observed when the two effects were combined.

Despite the increase in belief, the number of agents who actually performed the action fell 1.5% to be .54% of the eligible society. This was due partially due to the misinformation effect observed with the web site, in which agents learned the incorrect information from either of the interventions and thus did not learn sufficient information to take the action. The truncation effect of the literacy mechanism also played a role by preventing agents from learning sufficient information in order to take the action. Agents who were able to obtain incorrect belief information from the intervention were unable to obtain sufficient knowledge to change their behavior, leading to a decrease in the number of agents who were both knowledgable and who held the belief.

When the Internet access mechanism was activated, it only affected the web site but left the advertisement unaffected. As a result, many of the predictions of this model were similar to the results predicted for the web site. The web site predicted that knowledge in the society would decrease when Internet access was modeled, and this was indeed the case: knowledge decreased 1.3% in the society when Interent access was modeled relative to the combined case in which Internet acces was not modeled. The number of agents who believed the sceme was legal also increased relative to this baseline. The number of agents who believed the action was legal rose 4.7% relative to the least realistic case. In both of these cases, the lack of Internet access caused an increase in the number of agents who believed the behavior was legal because the intervention was less powerful and was not as effective at dissuading agents from performing the action.

While the number of agents who performed the action increased when Internet access was enabled for the web site alone, the number of agents who performed the action decreased by 0.6% when Internet access was enabled in the combined case. This decrease was somewhat surprising considering that the web site alone had similar trends in knowledge and belief but still lead to a net increase in the number who performed the action. In this case, however, there was a smaller increase in the number of agents who believed the action was legal due to the advertisement, since the agents who did lacked Internet access but otherwise were willing to communicate with an intervention could turn to the advertisement if they lacked access to the web page. Since such agents were more likely to be knowledgeable about how to perform the action, the addition of the advertisement could have a minimal overall impact on the society and yet have a more substantial impact on the agents at risk of performing the behavior should they believe the behavior was legal. This, in turn, could lead to the relatively sizable decrease observed between the two cases.

When newspaper readership was added to the combined case, it only affected the advertisement intervention. However, the effects of the readership change could be felt by multiple agents: the agents who were shut out from the advertisement intervention and could not learn knowledge or belief information; the agents who otherwise would have learned about the action and then gone to speak to the promoter; and agents who had increased similarity with an intervention and turned to the web page as opposed to the advertisement for information. In the combined intervention case with newspaper access, the last two points dominated. When readership was modeled, the number of agents who knew how to perform the action dropped 2.1% relative to the combined case in which no mechanisms were modeled. This was a dramatic change from the 1.2% increase observed when the advertisement was alone. Similarly, the 0.6% increase in belief between the combined case with newspaper readership and the combined case without readership was substantially smaller than the 1.1% increase observed between the advertisement with and without readership. Some of this decrease was due to less contact with the promoter, which lead to less knowledge and belief information in the society. However, the mean number of agents who communicated with the web page also picked up slightly, suggesting that the web site did compensate for some of the agents who could not speak to the advertisement and could provide additional illigality information.

In the readership case, the large decrease in the number of agents who knew how to perform the action was not offset by a corresponding increase in the number of agents who believed the action was legal. This lead to a substantial drop in behavior – 4.4% – the greatest observed in any of the interventions examined. While other cases had been associated with large drops in knowledge, such as either of the access cases, the corresponding increase in belief often lead to a greater percentage of the knowledgeable agents believing that the action was legal. In this situation, however, the decrease in knowledge was not offset by this corresponding increase in belief, a factor that contributed to the large decline observed.

When all three mechanisms were combined – the most realistic case for this intervention – the advertisement was affected by both the literacy mechanism and readership criteria and the web site was affected by the literacy mechanism and Internet access criteria. With all of these mechanisms active, the overall effect was a net 1.5% decrease in the number of agents who knew how to perform the action. Much of this effect was due to decreases in readership and access, which provided less information to the agents in the society. This decrease was further enhanced by the illiteracy mechanism, which helped to ensure that many of the agents who were able to interact with the intervention were unable to gain much information from it. This effect also played out in terms of belief: belief rose 4.5% between the most realistic case modeled and the combination case with no cognitive or access mechanisms modeled. Much of this increase can be attributed to the lack of interaction between agents in the society and the interventions. It is important to note, however, that the variances observed for these results were among the highest for all the variances observed for any of the cases. The fact that fewer agents had access to the interventions meant that there was a greater chance for outliers in terms of both knowledge and belief, making the mean a less reliable estimate of an outcome in the society.

The overall net change in behavior, however, was minimal. Despite the decrease in knowledge, there was a net increase in the number of agents who held the belief that the action was legal. After two hundred and fifty simulations, there was effectively no change in the number who changed their behavior: while a smaller number of agents knew how to perform the action, they were more likely to believe that the action was legal and that they should go ahead and perform it. The results observed align most closely with the results observed for Internet access, though the number performing the behavior were largely unchanged as compared to slightly decreased.

7.6 Experimental conclusions

In summary, the data presented above suggested:

- The effects of literacy and access mechanisms were small in absolute magnitude but could be noticeable in terms of relative effect. The increase observed when all literacy and information access parameters were enabled for the combined intervention case, relative to the combined intervention without any of the parameters enabled, was very close to zero. However, for the advertisement alone, the literacy and access mechanisms contributed to a 7.1% rise in the number who took action when comparing the least realistic case to the most realistic; for the web site alone, the change between the lest realistic and most realistic was 1.3%.
- The effects on knowledge or belief could be different than the effects observed for the number who changed their behavior. For instance, in the combination case, there was a decrease in the number of agents who knew how to perform the behavior and an increase in belief that the behavior was legal, yet the number of agents who changed their behavior was nearly unchanged. When the advertisement was modeled alone, the percentage change in knowledge was half the size that it was for the combined case and the percentage change in belief was about an eighth of what it was for the combined case, yet the number of agents who changed their behavior increased markedly. When the web site was modeled, the difference between the most and least realistic model were 2.1% and 6.8% in terms of knowledge and belief, respectively, although the increase in the number of agents who changed their behavior was only just above one percent. While some of this can be explained by the fact that different cognitive and access limitations had different effects on the society, several other factors were at work. Promoter contact rates, intervention contact rates, and intervention strengths were different for each of

the cases; additionally, the fact that knowledge and belief changed in certain ways did not necessarily mean that the same agent felt both (or even one) effect.

- Each of the different cognitive or access effects could have different effects on knowledge or belief. For instance, when the literacy mechanism was modeled, there was a decrease in belief for the advertisement, a big increase for the web site, and a moderate increase for the combined advertisement and web site. Effects did not necessarily have to be additive, however: literacy's effect on knowledge was to decrease it for the advertisement and web site relative to the individual intervention without literacy, but to increase knowledge for the combined case relative to the combined case without literacy.

- The different cognitive and access mechanisms can have differing effects on the standard deviations about the observed means. For instance, the access mechanisms for the isolated interventions have greater standard deviations than most of the other cases: the standard deviations for the advertisement is greatest when the newspaper is modeled, and the standard deviation for the web site is quite large when Internet access is modeled. While these effects are mitigated when all the access mechanisms are applied for each separate intervention, or when multiple interventions are active, it is generally the case that these access mechanisms lead to intervention results that are more likely to deviate from the mean than if the mechanisms were not modeled.

- Lastly, when all of the literacy, access, and readership parameters active, many of the same trends were observed as when they were not active [5]. For instance, the most realistic model of the advertisement lead to a slight decrease in the number of agents who knew how to perform the behavior while the web site lead to close to no change in knowledge; a similar general trend was observed in an earlier virtual experiment. The most realistic model of the combined intervention case had the lowest level of agents believing that the behavior was legal, followed by the web site and then the advertisement; a similar trend was also seen in earlier work. Lastly, while in the most realistic version of this experiment the web site and combination case ended up being almost equally effective, they were both more effective than the advertisement case.

It should be observed that the numbers obtained in the virtual experiment are averaged over multiple societies. The numbers given in tables 8, 9, and 10 are mean values, and the standard deviation values presented provides some guarantee as to the likelihood of differences from the predicted results. Construct, while a powerful modeling tool, is most effective for predicting aggregate outcomes or general behavioral trends, and may not be able to predict the outcome of a particular real-world condition with a high degree of accuracy. Much of this limitation comes from the difficulty of representing the observed world in a way that Construct can simulate, though some of it is due to the effect of un-modeled behaviors. Nevertheless, the aggregate results observed are likely to be observed in the world at large, though they may not be observed in any one specific data-gathering instance. While the impact of factors like illiteracy, lack of access, and lack of newspaper subscriptions may be important factors at the individual level – as such factors may lead some agents to change their behavior due to the fact that they cannot take advantage of an educational intervention – their overall aggregate effects on multiple simulated societies are seen to be relatively small.

8 Closing Comments

This technical report has discussed the addition of literacy and information access features to Construct. With these features, experimenters should now be able to assess the effects of literacy at the societal level or the aggregate group level, as well as the effects of Internet access and newspaper readership for individuals and groups. While the changes observed using these mechanisms are small in absolute magnitude, they can be large in magnitude relative to other changes.

Validation is an important aspect in any change to a modeling system, and several approaches will be taken to validate the modified model. First, the data presented in section 6 will be used when possible as such input reflects the real-world situation as well as can be obtained by the Construct development group. Additionally, several calibration experiments will be run to understand the effects of literacy and to determine the most appropriate values of the parameters discussed in section 5. However, a general validation model is extremely tricky for this data set. Since there are many other changes coupled with changes between literate and non-literate societies, such as changes in communication patterns, no ideal data set exists for validating the overall effect of the change. The lack of such a data set does not mean that validation should be ignored; instead, it suggests that alternative validation approaches should be sought in the future. Should such a data set become available, the model could be further validated at that time.

While this technical report has highlighted information related to literacy in Construct, it is important to note that the features described will eventually be expanded to a more general interface for different access properties. For instance,a more generalized version of the access formulas described in section 3 may be able to describe television access, radio access, and other access mechanisms for other forms of media. The misinformation effects may be suppressed for these characteristics, but the homophily, inference, and input properties till likely be identical. Such ideas may serve as the basis for continuing work with Construct in this domain.

References

[1] Kathleen Carley. A theory of group stability. *American Sociology Review*, 56(3):331–354, June 1991.

[2] Kathleen Carley. Smart agents and organizations of the future. *The Handbook of New Media*, 2002.

[3] Brian Hirshman and Kathleen Carley. Specifying agents in construct. Technical report, Carnegie Mellon University School of Computer Science, April 2007.

[4] National Assessment of Adult Literacy. A first look at the literacy of america's adults in the twenty-first century. *Institute of Education Sciences*, pages 1–28, 2006.

[5] Brian Hirshman and Kathleen Carley. Interactions in educational intervetions. *Sumbitted to IEEE SMC*, pages 1–20, 2008.

[6] Pew Internet. Pew internet and american life project. Web site, viewed 12/12/2007., 2007.

[7] Mediamark Research Inc. The daily and sunday newspaper audience report 2007. *Interactive Market Systems, Inc*, pages 1–31, March 2006.

[8] Kathleen Carley, Neal Altman, Boris Kaminsky, Demian Nave, and Alex Yahja. Biowar: a city-scale multi-agent network model of weaponized biological attacks, 2004.

[9] Craig Schreiber, Siddhartha Singh, and Kathleen Carley. Construct - a multi-agent network model for the co-evolution of agents and socio-cultural environments. Technical report, Carnegie Mellon University School of Computer Science, May 2004.

[10] Stuart Russell and Peter Norvig. *Artificial Intelligence: a modern approach*. Prentice-Hall, Upper Saddle River, NJ, 2nd edition, December 2002.

[11] Christopher Bishop. *Pattern Recognition and Machine Learning*. Springer, Singapore, Singapore, 1st edition, August 2006.

[12] Jonathan Kozol. *Illiterate America*. Doubleday & Company, New York, NY, 1985.

[13] National Center for Statistics. Adult literacy in america. pages 1–201, April 2002.

[14] Irwin Kirsch. The international adult literacy survey (ials): understanding what was measured, December 2001.

[15] Pew Internet. Internet activities. December 2007.

[16] Gretchen Livingston. *Latinos Online*, pages 1–23. Pew Internet & American Life Project, 2007.

[17] Newspaper Association of America. About the newspaper association of america, December 2007.

[18] Newspaper Association of America. Total paid circulation, circulation by volume. December 2007.

[19] Mediamark Research Inc. Newspaper daily section readership 2007. *Interactive Market Systems, Inc*, pages 1–36, August 2006.

[20] Brian Hirshman and Kathleen Carley. Information access experiments using construct. Technical report, Carnegie Mellon University School of Computer Science, March 2008.

[21] Charlie Catlett et al. Teragrid: Analysis of organization, system architecture, and middleware enabling new types of applications. *IOS Press*, 2007.