The Impact of Educational Interventions by Socio-Demographic Attribute

Brian R. Hirshman, Michael Martin,
Michael W. Bigrigg and Kathleen M. Carley

July 10, 2008
CMU-ISR-08-118

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

Center for the Computational Analysis of Social and Organizational Systems
CASOS technical report.

This work was supported in part by the IRS project in Computational Modeling and the NSF IGERT in CASOS (DGE 997276) and by the National Science Foundation through TeraGrid resources provided by Purdue University. 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 Air Force 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.

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

Abstract

Past work with the Construct simulation tool has asked questions at the population level: what percentage of the total population knew a fact, held a belief, or performed an action. Since several cognitive and access mechanisms have been added to Construct, and these attributes have been tied to socio-demographic sub-populations, it is now possible to examine the effects of educational interventions on sub-populations. This technical report presents a virtual experiment which analyzes a simulated population by various socio-demographic attributes. It specifies the experimental design, describes parameters used, presents a series of results which explore the effect of literacy and information access on societal sub-populations, and identifies educational interventions which may and may not be effective.

Table of Contents

List of Figures

List of Tables

1 Introduction & Motivation

Recent success in modeling a variety of interventions has lead to follow-up questions about information access. For instance, if socio-demographic traits are commonly correlated with illiteracy, will a print intervention be effective in modifying their behavior? Alternatively, if a high-risk population lacks Internet access, will a sophisticated web site be effective? Construct, a multi-agent simulation platform developed by the CASOS center, has previously been used to ask questions about the relative strengths of various educational interventions [1][2]. In order to understand the effects of these physical and cognitive barriers to information access, a variety of virtual experiments were conducted using Construct to understand the magnitude of these effects. This technical report describes the setup for these experiments, presents the results obtained from the simulation, and discusses the effect of the interventions by socio-demographic attributes.

The remainder of this technical report is organized as follows. Section 2 outlines the virtual experiment – the problem being modeled, the types of societal features modified, the information access parameters used, and other important global parameter settings. Section 3 discusses the kinds of modifications made for each virtual experiment, describing the interventions and information access mechanisms varied over the course of the runs. Section 4 describes the socio-demographic distributions of the virtual experiments performed. Section 5 presents the aggregate experimental result, while Sections 6 through 8 present experimental results broken down by socio-demographic characteristics. Section 9 highlights several overall trends, and Section 10 concludes.

2 Experiment Parameters

2.1 Construct Overview

Construct is a social network analysis tool which examines the evolution of networks [3][4][5]. Specifically, it models the processes by which information moves around a social network. Construct seeks to combine social network analysis, a field that has typically sought to describe static networks, with an understanding of information diffusion to create a simulation environment that models network change over time. Construct is an agent-based model, which means that individual actors are the sources of decisions in the model. Agents in Construct interact via homophily – the principle of “like attracts like” – which leads similar agents to interact, share knowledge, and become more similar to each other. As a simulation engine, Construct draws from a wide variety of fields: social network research for analytic techniques, sociology for its interaction mechanisms, psychology for agent choices and decision procedures, computer science for its algorithms and artificial intelligence as the glue that binds the various ideas together. Past work with Construct has sought to compare and contrast different interventions on a social network, to examine belief diffusion and information propagation among people, to study the effect of different network topologies on diffusion rates, and several other topics [1] [2].

Additional information about the Construct system is available in the references of this paper. The Construct system itself is freely downloadable from the CASOS website, http://www.casos.cs.cmu.edu/ projects/construct.

Table 1: Numbers and Weights on Key Facts

Fact Type		Characteristics		
		number	interaction weight	transmission weight
Action exists	knowledge	1	1x	3x
How to perform action	knowledge	6	1x	3x
Action is right	belief	3	1x	3x
Action not right	belief	4	1x	3x
Action is worth doing	belief	3	1x	3x
Action not worth doing	belief	3	1x	3x
General knowledge	other	500	1x	1x

2.2 Modeled Scenario

The overall suite of experiments run, of which this experiment was just one part, sought to model the effects of different cognitive and access constraints on the number of individuals who could interact with an educational intervention [1][6]. The educational intervention sought to encourage agents to perform in a specific way; in the taxpaying domain in which this problem was conceived, the interventions can be thought of as seeking to dissuade individuals from participating in an illegal tax scheme, to encourage them to take a legitimate tax credit, or to help them their file tax forms in a correct fashion. One of the motivating features for the project was modeling behavior that encouraged agents to legitimately take the Earned Income Tax Credit (EITC), but the conceptual framework can be generalized to other sorts of educational interventions [7]. The population modeled in the society was chosen to be predominantly low income, since this population was known to be more likely to suffer from lack of information access [8][9][10]. By using societies comprising of largely low-income agents, the literacy and access mechanism modifications could have a more pronounced impact as compared to other types of agent populations.

2.3 Knowledge and Belief

As has been done in previous work, the action examined was modeled as an action to be deterred. The action was modeled using two components: "knowledge" and "belief" [1][2]. Knowledge represented the facts necessary to perform the action; without sufficient knowledge, agents would be ignorant of the action and would not perform it even if they believed that they should. Belief was a function of three parts: a sequence of facts, similar to the knowledge component described previously; the belief an agent held in the previous time period; and a weighted function of the beliefs of agents in the surrounding social network. Table 1 presents the types of facts associated with knowledge and belief, the number of facts associated with each, and their interaction and transmission weights.

Knowledge was further subdivided into existance and performance facts. A single fact was used to represent whether the agents knew that the action existed. If an agent knew this fact, then that agent was said to "know of" the action. This was done to model the spread of the action in the society, and was given a higher weight for transmission between the promoters, interventions, and other agents. Agents who knew this fact could perform the action; all agents lacking this fact were unable to perform the action. By design, many more agents could know that the action existed even if they did not know how to take the action, as is often observed in real-word social networks.

The other type of knowledge facts were facts about how to perform the action. As can be seen in the first column of Table 1, six facts were used to represent whether the agents knew enough information about the action in order to perform it. In order to successfully perform the action, an agent needed to know at least three of these six "know how" facts. Any three of the six facts were sufficient. This setup allowed the experiment to model how two agents could perform the action using slightly different understandings, a scenario commonly occurring in the real world. This setup also ensured that agents could have some variability in the composition of their knowledge when making the decision. All six of the facts were weighted equally; an agent was said to perform the action just as effectively regardless of whether it the first three facts, first two facts and the fourth fact, or the last three facts.

The second requirement for taking the action was to have sufficient belief. There were two beliefs in the model: belief that the action was legal and therefore a legitimate thing to do, and belief that the action was a good idea to do (often written that the agent "should do" the action) regardless of its legality. Agents held a belief if sum of their current knowledge, prior beliefs, and social influence was above a certain threshold. To take the action, an agent only needed to be above this threshold; any additional belief beyond this threshold was superfluous.

The legality belief was modeled with seven facts, three facts associated with the idea that the action was legal and four with the idea that the action was not legal. Each fact was weighted equally, so the belief knowledge component was a linear function of the number of facts known related to pro or con belief. Thus, the setup of three pro-action and four anti-action belief facts suggested that the action would be illegal to an agent with full information. This also helped to guarantee that illegality beliefs would be passed slightly more frequently than legality beliefs. When interacting, agents would be more likely to communicate a fact suggesting that the action was illegal in comparison to a fact suggesting that the action was legal, since it was likely that they knew more of them.

The "should do" belief was modeled with six facts, three of which suggested that performing the action was a good idea and three of which suggested that performing the action was a bad idea. Again, each fact was weighted equally so the belief knowledge component was a linear function of the number of known belief fact. Unlike the legality belief, there was no imbalance between the number of facts associated with each belief. This meant that, if an agent knew all the facts associated with the belief, it would not necessarily be predisposed against the action.

For an agent to hold a belief, it would often need to know more facts relating to one of the belief statements than it knew for the other. For instance, agents who knew one legality facts but zero illegality fact would believe the action was legal. If the agent learned two illegality facts, it would hold the legality belief for a time, since the influence of the past belief would linger and the agent's belief would not be modified instantaneously. After a short time, however, the agent would believe that the action was illegal because the two illegality facts were greater than the one legality fact. These effects could be temped by social influence, which might have a large or small effect on the agent depending both on the influentialness of one agent and the influenceability of the other.

In order to perform the action, an agent had to either believe that the action was legal or that the action was a good idea. If the agent held the legality belief but not the should-do belief, this represented a case where the agent believed that the agent did not mean to be deceptive but was genuinely misinformed as to the legality of the action. If the legality belief was not held but the

Table 2: Initial Distribution of Key Facts

Fact Type	Chance that individual fact is known by		
	human agent	promoter	seminar
Action exists (1 fact)	0%	100%	100%
Action know-how (6 facts)	0%	100%	100%
Believe right (3 facts)	1%	100%	100%
Believe not right (4 facts)	5%	0%	0%
Believe worth doing (3 facts)	1%	100%	100%
Believe not worth doing (3 facts)	5%	0%	0%
General knowledge (500 facts)	20%	2%	2%

should-do belief was, this represented a case in which the agent intentionally attempted to commit fraud. Agents who did not believe the behavior was legal and who did not believe that it was a good idea – by far the majority group in the simulated population – were unable to perform the action, even if they had sufficient knowledge.

Additionally, a total of five hundred "social facts" were used in the experiment, facts which had no bearing on whether agents performed the action. These facts served to increase or decrease the homophily between two agents and helped agents select interaction partners. Due to the preponderance of these general knowledge facts, they were exchanged in the vast majority of interactions. As the simulation progressed, agents would exchange these facts, gradually becoming more similar to some agents and (relatively speaking) less similar to others. While the social facts did not directly affect the number of agents who performed the action, they indirectly helped to define the interaction partners of an agent and therefore greatly influenced agent sources of information.

2.4 Simulated Society

The experimental society was comprised of three thousand human agents for all three of the virtual experiments performed. While the initial knowledge holdings of an individual agent depended on the experiment, agent knowledge facts were drawn from the underlying distribution as described in Table 2.

The first column of Table 2 describes the knowledge associated with the human agents. The human agents did not have any of the knowledge facts initially, so they had to learn these facts via interaction with an outside source such as a promoter. Human agents, however, did have initial beliefs. Agents have a 5% chance of knowing each of the anti-action facts (that the action was illegal and that they should not do the action), and a 1% chance of knowing each of the pro-action facts (that the action was legal and that they should do the action). This lead to a society in which a small minority of agents initially held pro-action beliefs, a larger number held an opposing belief, and the majority of the agents did not have any belief at all and were primarily swayed by the beliefs of agents comprising their surrounding social network.

Additionally, the human agents had a 20% chance of knowing each of the five hundred social facts, meaning that each of the agents initially knew about a hundred social facts in order to jumpstart agent-agent differentiation and homophily. These facts were selected from a random uniform distribution. Due to small perturbations between the types of facts known by each agent, agents were more similar to some agents than to others. These small differences would influence interaction, and would help determine interaction patterns.

The experiment was designed so that the action was propagated through the society by a single "promoter" agent, an agent who entered the society at the beginning of the first simulated period and attempted to convince as many human agents as possible to perform the action. The promoter had an associated companion, a "seminar" agent, which was active less frequently but could deliver a more powerful message to its attendees. This allowed the promoter to, from time to time, deliver a more powerful message to the human agents who chose pay attention to it.

Both the promoter and the seminar had full knowledge of the action and how to perform the action. When there was no intervention present, they served as the sole source for all knowledge about the action. The promoter and the seminar knew all the is-legal facts and the should-do facts in an attempt to convince as many agents as possible to perform the action.

The simulation ran for one hundred and four time periods. This represented one year with two time periods per week for the simulated society. Human agents were active every time period, choosing one potential interaction partner to communicate with and receiving communication from one another agent. The promoter was active only every other period, but could initiate communicate with several agents and receive communication during that time. The seminar was active only four times during the entire simulation, could not contact other agents, but could receive communication from a larger number of agents than the promoter was able.

2.5 Network Topology

The human agents in the virtual experiment were all linked together via an underlying social network. This allowed one agent to interact with a subset of the rest of the agents in the network but not with all of the possible agents. However, these subsets were not all alike. Some slices of the agent population could form small, highly interconnected cliques, while other slices could have links that spanned different parts of the networks, while still others could have a jumble of assorted edges. The underlying network topology helped to define this type of structure, though it was up to the individual agents to interact according to homophily and build an actual communication structure on top of the possible communication framework provided by the network topology.

The density of the human agent to human agent social network was about .0133. This meant that each of the three thousand human agents had about forty neighbors. An independent social network was created for each replication, as drawn from the underlying distribution of social network topologies, to ensure that the results observed were not biased by the specific layout of any one individual social network.

The network topology used in this research was a uniform random graph [11]. The uniform random network was a random network of three thousand nodes, created by selecting 1.33% of the possible edges from the fully-connected graph and creating the social network from them. The uniform random network was generated in a manner consistent with the generator suggested by Erdos and Renyi: edges were selected randomly from a complete graph until the requisite density was achieved [11]. No preferential attachment, triadic closure, or clustering algorithm was used in this process, but the edges were selected to ensure symmetry (if an ego could contact an alter then the alter could contact the ego). This process created a network in which the mean number of neighbors was forty, the average distance was empirically determined to be 2.5, and the diameter was three. While it has been shown that Erdos-Renyi graphs do not capture some of the subtleties in real-world social networks, they provide a reasonably accurate first approximation of them.

While it was not performed for this particular work, past research using Construct has also investigated different types of underlying social networks. For instance, the paper "Societal Factors as Moderators of Intervention Strength," a paper draft submitted to IEEE SMC, suggests that different underlying social networks may lead to different observed results [1]. Much of the work described there was used in this study, including the descriptions of the simulation setup, knowledge distribution descriptions, and interaction sphere descriptions. One noticeable change, however, is the fact that the number of agents in this experiment has been increased to three thousand (from two thousand as described in that paper); this has resulted in scaling of other simulation parameters.

2.6 Other Parameters

Though facts and beliefs were introduced in Section 2.3, there are a number of subtleties worth mentioning with respect to how these facts and beliefs were used.

For instance, agent did not use the knowledge holdings of another agent when computing a homophily score with another agent. Instead, when an agent attempted to compute its similarity with another agent, it would use a feature called "transactive memory" to compute its similarity score with another agent. Transactive memory, represented second-order information about a fact; transactive memory could represent that an ego knew that an alter knew a specific fact [4][12]. Thus, agents computed knowledge similarity using their perception of the agent's knowledge, not the agent's knowledge itself. Similarly, agents used a special form of belief, belief transactive memory, when computing the belief of the surrounding community; unlike fact transactive memory, which was binary (either an agent knew it or it did not), belief transactive memory was real-valued and therefore more nuanced. In this way, transactive memory served as a mental model of another agent's knowledge and can have many of the features seen in a realistic mental model. The transactive memory perception could be inaccurate, incomplete, or incorrect, depending on the situation – alter agents could learn, evolve, or change beliefs without an ego agent knowing about it. An ego's transactive memory of an alter agent was only updated if an agent sent a specific fact to the alter agent, or if the ego agent learned (possibly from a third party) that the alter knew a particular fact. About twenty percent of all messages sent between agents contained exclusively transactive memory to help keep this information up-to-date.

Additionally, though knowledge and beliefs were both treated as pieces of information that could be exchanged between agents, knowledge and belief were represented slightly differently in the simulation. For example, knowledge was represented as binary facts: either the agent knew a fact, or the agent did not. Agents could not partially learn a fact, nor could they partially tell the fact to another agent; communication was all or none. Forgetting was not enabled, so agents retained any information that they learned over the course of the experiment. In contrast, belief was allowed to vary between agreement, disagreement, and no information along a real-number scale. An ego agent could modify its belief by learning new information from an alter, or by finding out that several alter agents in their social network had changed their beliefs, leading to a change in the prevailing societal belief. However, belief was binarized when determining whether an agent would perform an action. At the end of the simulation, a hard cutoff was applied to belief, effectively turning the real-valued belief into a binary yes-or-no decision.

Lastly, facts were not cognitively equivalent to the agents in the simulation. As can be seen in the second and third columns of Table 1 on page 2, agents had different interaction and

Table 3: Distribution of Key Facts for Intervention

Fact Type	Intervention				
	Ad campaign	Web site	Call center	Radio ad	Mailing
Action exists (1 fact)	100%	100%	100%	100%	100%
Action know-how (6 facts)	10%	33%	10%	10%	10%
Believe right (3 facts)	0%	0%	0%	0%	0%
Believe not right (4 facts)	33%	100%	100%	33%	33%
Believe worth doing (3 facts)	0%	0%	0%	0%	0%
Believe not worth doing (3 facts)	33%	100%	100%	33%	33%
General knowledge (500 facts)	10%	2%	5%	10%	10%

transmission weights for particular facts. In the experiment described, knowledge facts, belief facts, and general facts all had an interaction weight of one. This meant that agents considered all facts equally when evaluating a potential interaction partner's knowledge. Since agents often held vastly more general knowledge facts than they did knowledge facts, this lead to situations in which human agents were selecting interaction partners primarily based upon their general knowledge similarity – the knowledge played only a minor role. However, the fact transmission weight for action-related facts was much higher: in order to over-sample for conversations of interest for facts related to the action, the fact transmission weight was set to three. This ensured that agents were more likely to pass information related to the action in a given interaction (if they knew it), but the passing of this type of information was still relatively infrequent in the general society. The net effect of these weights was to ensure that general knowledge homophily largely drove agent-agent interaction, but even when knowledge homophily was dominant agents would be more likely to transmit action-related facts.

3 Virtual Experiment Setup

In the experiments performed, two classes of parameters were varied. The first parameter varied was the number and type of interventions present in the society, discussed in Section 3.1. These represented the attempt of an outside entity attempting to influence the behavior in a society by a series of educational interventions to warn at-risk agents about the action. The second parameter varied was the information access method that was active and is further explained in Section 3.2. These access methods could weaken the intervention in various fashions, by restricting the number of agents who could interact with the intervention or cutting down the amount of information learned from the intervention if interaction occurred. The experiment introduced in Sections 4, and whose results are discussed in the remainder of this document, manipulates these parameters.

3.1 Interventions

In addition to the society of human agents and the promoter, there were also up to five other intervention agents. These interventions were representations of print advertisements, web pages, call centers, radio advertisements, and postal mailings. Interventions had differing amounts of knowledge and belief facts; however, all were similar in the sense that they could not initiate communication with a human agent and had to be contacted by an informed party before spreading their information. The knowledge holdings of the individual intervention agents are highlighted in Table 3. The fraction of agents who could communicate with the interventions is highlighted in Table 4.

Table 4: Percentage of Population That Could Interact With an Intervention

Property	**Intervention**				
	Ad campaign	**Web site**	**Call center**	**Radio ad**	**Mailing**
Percent of society	*	*	50%	90%	50%

* affected by the access mechanisms of Section 3.2.

- The advertisement was meant to provide a small amount of knowledge and belief, while also containing a large amount of general knowledge information to make agents more likely to examine it. This meant that the intervention could have a small impact on a variety of agents, and could serve to dissuade agents before they ever learned about the action. The advertisement had the action exists bit, had a ten percent chance of knowing each of the six know-how facts, and had a thirty percent chance of knowing each of the anti-action belief facts (for both legality and agent should do). It could only send a short message, but it had about ten percent of all the social facts in the society – the highest of any of the interventions. The number of agents who could interact with the advertisement was dependent on newspaper readership when active; when it was inactive, it was set to 100%.

- The web site was designed to provide a large amount of belief information by proving a large number of reasons why the action was illegal and a bad idea. In doing so, however, it could potentially be scraped for knowledge information. The web site knew the action existed and had a thirty percent chance of knowing each of the knowledge facts necessary to complete the action. It knew all of the anti-action belief facts in order to convey a long, strong anti-behavior message. The interaction sphere of the web site was dependent on the Internet access parameter as described in Section 2.5; when it was inactive it was set to 100% as a comparison case.

- The call center was designed to answer agent questions with regard to the action. It knew less information about the action than the web site, but had the same percentage of the belief facts. It had more social knowledge then the web site, meaning that agents were more likely to communicate with it than with the web site. However, it could not as much information one interaction as the web, and also was not accessible by all the agents in the simulation.

- The radio advertisement was very much like the print advertisement: it was designed to provide a small amount of information and belief while reaching a large number of agents. It too could only send a short message. Unlike the advertisement, it was not affected by the literacy and access mechanisms of Section 3.2.

- The postal mailing was designed to represent an official mail intervention conveying information about the action with the intent to deter at-risk agents from performing the behavior. It had the same information content as the advertisement, but it had a different interaction mechanism than the other interventions. When the mailing agent was active, it could send a message to some fraction of the 50% of the agents in the society. For the next four time periods, the mail message resided in the agent's “mailbox”. The agent then had a certain probability of checking their mail and learning the information in the mailing. Agents who read the mailing absorbed some of the information contained in the mail message; agents who did not were unaffected by the intervention.

These five intervention types lead to the development of twelve intervention cases used in the experiment. The first case, a baseline case, was a case with none of the interventions present. This was used to measure the amount of activity occurring if the promoter was allowed to run its course unimpeded. The next five cases were single-intervention cases: the advertisement alone, the web site alone, the call center alone, the radio advertisement alone, and the mailing alone. Six combination cases were also examined. Three of these cases were combinations with the web site: the advertisement and the web site, the radio advertisement and the web site, and the mailing and the web site (the call center and web site intervention was explored in previous work). The remaining three were super-bundles containing more complex interventions - the mailing, radio ad, and web site; the call center, mailing, radio ad, and web site; and the print ad, call center, mailing, radio ad, and web site (all of the interventions together).

3.2 Information Access

In addition to the intervention modifications, three different types of information access mechanisms were used: literacy, Internet access, and newspaper readership. These access mechanisms affected whether agents could interact with a specific intervention (if the intervention was present).

- The literacy mechanism affected all interventions that required reading printed material. This meant that the print advertisement, the web site, and the postal mailing were affected. When this information access parameter was enabled, illiterate agents could still access these interventions; however, the agents did not receive the full message (and could even suffer from message distortions). Literate agents were unaffected.

- The internet access mechanism affected all interventions that required web access, namely the web page. When this information access parameter was enabled, agents lacking web access could not communicate with the web site intervention at all. Agents with Internet access were unaffected.

- The readership mechanism affected all interventions that required newspaper access, namely the print advertisement (which was assumed to appear in a print newspaper or associated magazine). When this information access parameter was enabled, agents lacking newspaper subscriptions could not communicate with the advertisement intervention at all. Agents with subscriptions were unaffected.

To determine which agents were literate, had Internet access, or obtained a newspaper or a regular basis, data was gathered from a number of places including the National Center for Statistics, the Pew Research center, and Newspaper Association of America [8][10]. Results were gathered for aggregate population groups, including breakdowns by gender, age, race, education, and income. The changes necessary to support these interventions are described in the technical report “Modeling Information Access in Construct” [6].

The virtual experiments in this research used five different levels of information access. The baseline case had all of the mechanisms disabled for comparison with earlier experiments. Three cases enabled each of the mechanisms separately to try and understand the individual effect of

Table 5: Agent Breakdown for VE2 Stylized City

Race		Education	
White	*	Less than high school	40%
Black	*	High school diploma	30%
Asian	*	Bachelors degree	29%
Hispanic	*	Grad / prof. school	1%
Other	*		
Income		**Age**	
$0-15k	40%	0-30	20%
$15k-30k	35%	30-40	20%
$30k-50k	15%	40-50	20%
$50k-80k	6%	50-60	20%
$80k-120k	3%	60+	20%
$120k+	1%		
Gender		**Parent**	
Male	50%	Yes	50%
Female	50%	No	50%

* 100% for one race, 0% for all others in each replication

each of these mechanisms. Finally, a combination case enabled all three access mechanisms, thereby attempting to model processes in the real world as realistically as possible.

4 Virtual Experiment II: Stylized Cities by Ethnicity

The virtual experiment performed examined several types of stylized cities in order to tease out the various effects of agents who performed the action. Five variations on a stylized city were replicated multiple times in order to create an agent database from which to sample for the number and types of agents which performed the action, who knew sufficient information to perform the action, and who believed that the action was legal.

4.1 Agents Used

The stylized city used in this experiment was similar to other stylized cities used in past virtual experiments [6]. In this experiment, however, there were minor modifications to the experiment design. While agents were over-sampled to ensure that a high number of low-income and low-education agents were represented in the society – the groups that were most likely to suffer from illiteracy and a lack of Internet access – a decision was made to perform a virtual experiment separately for agents of each race. This experimental design is presented in tabular form in Table 5, and graphically in Figure 1.

Rather than making assumptions about the racial composition of the stylized city, these runs were performed by assuming that all the members of the society were of a single race. For this virtual experiment, then, five sets of runs were performed: one each in which all the agents in the city were treated as White, Black, Asian, Hispanic, and Other. Thus, the starred values in the above table became 100% for the agent set of interest and 0% for the remaining four groups. While this assumption does not necessarily hold in the outside world, it made it possible to perform tests on the other parameters which might otherwise be confounded with race. For instance, research has suggested that the number of individuals with internet access differs substantially by race [9][13]. By eliminating this effect within a particular run, it was hoped to

Table 6: Modifications Active in the Most, Least Realistic Cases

Intervention name	active modifications	notes
No intervention	none	
Advertisement	literacy, readership	
Call center	none	
Mailing	literacy	
Radio message	none	
Web site	literacy, internet access	
Ad + web	all three	
Mail + web	literacy, internet access	
Radio + web	literacy, internet access	
Mail + radio + web	literacy, internet access	
Call + mail + radio + web	literacy, internet access	"all_but_ad"" in graph labels
Ad + call + mail + radio + web	all three	"all"" in graph labels

remove the variance that might occur within an individual run, as modeling race could lead to differences in information access rates among agents in the same experiment. By modeling race in this way, race did contribute to variance when all the runs of a certain type were aggregated; however, race did not contribute to heterogeneous non-local effects on agent communication patterns within a run.

In the virtual experiment, the agent population was skewed predominantly toward low income individuals. To this end, the promoter and the interventions were given increased homophily with agents at the lowest income level (either $0-$15k or $15k-$30k) and not those with professional degrees (less than high school, high school diploma, or bachelor's degree). These agents comprised the vast majority of the society and were the primary target for the behavior. When computing statistics on the society, however, only a subset of the agent population was used. In discussions with simulation experts, it was decided that if the agent was not in the income range or not in the education range, it would have a 50% chance of being able to performing the action, even if they knew enough about it and believed the action was legal. If the agent did not match the income level or the education level, it would only have a 25% chance of being able to perform it. Agents who did not match, and therefore could not perform the action, were modeled as those likely to be flagged and caught if they performed the action. While this flag prevented the agents from performing the action directly, it did not prevent them from serve as conductors of information throughout the society; these agents could pass on any knowledge or belief fact in their communication with other agents in the society and could even communicate with the promoter if they desired.

Using these matching criteria, slightly less than 75% of the agents in the simulated society had the necessary characteristics to perform the action. Since the other 25% of the population could not perform the action, they were excluded from the analysis. All resulting percentages presented represent statistics gathered on these matching agents, even in the socio-demographic breakdowns related to income and education.

4.2 Experiment Design

The experiment cases were first generated as a 12x5 matrix, with twelve interventions crossed by five races, for a total of sixty experiments. Several additional runs were also performed to examine the effects of the literacy, internet access, and newspaper readership experiments individually; this increased the total number of experiments run by a factor of about three. A total of just over two hundred experimental cases were eventually submitted to the Condor cluster at the TeraGrid supercomputing center in West Lafayette, Indiana [14]. Each case was replicated fifty times; in all, a total of 10,000 individual instances of the Construct simulation were performed.

The experiment was performed in two days using the heterogeneous cluster of computers available on the TeraGrid [14]. Since the TeraGrid had a very large number of machines, replications could be performed in parallel and computation could proceed rapidly. For comparison, an earlier instance of the experiment was performed on machines local to the CASOS center, including two dual-core, 4CPU machines with 64GB of memory. These 10,000 experiments took almost a month to perform.

Table 6 summarizes the literacy and information access modifications used in the model, the one in which all of the applicable information and access parameters were active. In these most realistic cases, the literacy mechanism was used any time there was a print advertisement intervention, a web site intervention, or a mail intervention; the newspaper readership access modification was used when the print advertisement was active; and the internet access mechanism was used when the web site was active. In some intervention combinations, all three of these mechanisms were active simultaneously, such as occurred in the most realistic model of the advertisement and web site. Only the relevant interventions were employed in the combined cases: while literacy might affect both the ad and the web site, the newspaper readership modification affected only the advertisement. In three of the cases – the no intervention case, the call center alone case, and the radio alone case – none of the mechanisms were modeled since they were not relevant.

5 Overview of Results

5.1 Tables and Figures

Two sets of tables and one set of figures were used to summarize the virtual experiment. The tables, such as Table 8 (pages 15), display the three most effective interventions for the parameter in question, as well as the single least effective intervention. For each of these four interventions, the following values were recorded.

- The name of the intervention or intervention bundle. This was one of the intervention bundles described in Section 3.1: the no-intervention case (none), the print advertisement (ad), the web page (web), the call center (call), the radio message (radio), the mail message (mail), the ad and web combo (ad+web), the mail and web combo (mail+web), the radio and web combo (radio+web), the mail-radio-web combo (mail+radio+web), the mail-call-radio-web combo (all but ad), or the ad-mail-call-radio-web combo (all).

- The value MIR_{kji}, the mean percent of matching agents who did k (either knew enough information to act on the behavior, believed the behavior was legal, or actually took action, depending upon the table) for socio-demographic parameter j when intervention i was present and the most realistic model of agent ability and access to information was used. This value was a mean of means, as it represented the average of individual percentages computed for each run. This provided a measurement of the expected outcome.

Table 7: Most Effective Interventions for the Aggregate Society

Case	metric knowledge	 belief	 Behavior
Fewest knowing how	all	all	all
(% who knew how)	3.15%	3.02%	0.527%
(standard deviation)	0.52%	0.48%	0.145%
Second	radio + web	all but ad	radio + web
(% who knew how)	3.15%	3.08%	0.536%
(standard deviation)	0.59%	0.53%	0.156%
Third	web	radio + web	web
(% who knew how)	3.16%	3.14%	0.553%
(standard deviation)	0.53%	0.50%	0.146%
Least effective	mail + radio + web	call	mail
(% who knew how)	3.26%	3.56%	0.601%
(standard deviation)	0.61%	0.59%	0.165%

- The value VIU_{kji} was the variance around the mean percent of who did k (either knew enough information to act on the behavior, believed the behavior was legal, or actually took action) for socio-demographic parameter j when intervention i was present and the least realistic model of agent ability and access to information is used. This provided a measurement of how much individual variability would be likely to affect a particular outcome.

A second set of tables, such as Table 11 (page 19) compare the stylized cities examined in this document with percentage changes with the data modeled for Construct replications performed on several real cities: Hartford, CT; San Diego, CA; Orlando, FL; and Kansas City, MO. These tables contain the names of the most effective interventions for reducing knowledge, belief, and behavior. Additionally, the fraction of the population which met the socio-demographic criteria is also noted. For a more detailed description of the real cities, additional information is available in the technical report "The Impact of Educational Interventions on Real and Stylized Cities" [2].

A set of figures included in Appendix B through Appendix D (page 37 onward) present the relative effects of each of the interventions. In these figures, each of the eleven educational interventions and intervention bundles are presented as relative percentages to the no-intervention case. Interventions that had the same mean values as the no-intervention case were represented as no change, while values that were lower than the no-intervention case were represented as decreases and drawn as bars pointing left. These figures were created as clustered bar charts, with each of the socio-demographic parameters clustered together by intervention for clearer reading.

5.2 General Effects

Before delving into the specific effects of the socio-demographic parameters on interventions, it is necessary to understand the effects of the interventions on the overall city. These results are presented in Table 7. The stylized city examined in this set of virtual experiments differed slightly from past work with stylized cities, including the stylized city experiments examined in the technical report "The Impact of Educational Interventions on Real and Stylized Cities" [2]. In that set of experiments, the cities used were heterogeneous by race. Since race was correlated with several of the literacy and information access parameters examined, it was not possible to fully explore the effect of literacy and access parameters on sub-populations [6]. The number of replications performed in that case was too small in order to separate out the effects of any of the other socio-demographic parameters from race, so the experiments were re-run in order to separate out race as a factor. The experiment results described in the remainder of this experiment – and in the remainder of this technical report – thus differ slightly from what was seen in the heterogeneous city environment.

When an ANOVA was run for knowledge, no statistically significant effect was found ($p < .69$). The difference between the most effective intervention at reducing knowledge, the all-intervention case at 3.15% and the mail+radio+web at 3.26% was only a small fraction of the variance observed over several iterations of the simulation. While the three most effective and one least effective interventions have been provided in Table 7, the differences between the all-intervention combination, the radio+web intervention combination, and the web intervention alone are all very slight.

An ANOVA run for belief suggested a highly significant effect for belief in the overall society ($p < .0001$). The all-intervention combination lead to the greatest decrease – only 3.02% of the agents held the belief – while the call center intervention lead to a slight increase in belief relative to the no-intervention combination. The all-but-ad intervention and the radio+web intervention also had important impacts on belief, since these intervention combinations could spread belief information to a large number of agents and thereby decrease the number of agents who either started out with a particular belief or were directly or indirectly swayed by the promoter.

A statistically significant effect ($p < .001$) was found when examining whether intervention affects behavior. For the society as a whole, the all-intervention combination was found to be the most effective intervention at reducing behavior, followed by the radio+web intervention, followed by the web intervention alone. However, the mail intervention was one of four interventions that were identified as having a statistically significant increase in the number of agents who knew enough information to perform the activity as compared to the no-intervention case. This finding suggests that the all-intervention combination and the radio+web combination were effective due to the combined effect that it could have on knowledge and belief. By decreasing the number of knowledgeable agents as well as the number of agents who held a particular belief, they were able to contribute to a slight decrease in behavior in the society.

It is important to note, however, that only a relatively small percentage of the society had both sufficient knowledge and belief – only about one agent in two hundred, a mean value of 0.566%, actually made a decision to do so – so the number of agents affected was very small generally. Nevertheless, this represented a fairly significant fraction of the knowledgeable agents in the society as well as a substantial fraction of the belief-holding agents in the society.

6 Specific Effects by Age

Age was one of the most important socio-demographic attributes which contributed to how agents obtained information. As discussed in Section 4.1, five age categories were used: 0-29, 30-40, 40-50, 50-60, and older than 60. These age ranges were selected as generic age goups, and individual agent ages were not assigned for the purpose of the simulation. Agents were more likely to communicate with agents in their age group, but no special preference was made for agents in one age range to communicate with those in the neighboring age range: agents in

Table 8: Greatest Changes in Knowledge, by Age

Case	agent age range				
	0-29	**30-40**	**40-50**	**50-60**	**60+**
Fewest knowing how	ad + web	mail + web	all but ad	ad	all
(% who knew how)	3.04%	3.13%	3.09%	2.99%	3.07%
(standard deviation)	0.42%	0.40%	0.43%	0.38%	0.40%
Second	all but ad	radio + web	web	all	radio + web
(% who knew how)	3.10%	3.14%	3.14%	3.07%	3.10%
(standard deviation)	0.39%	0.48%	0.34%	0.41%	0.42%
Third	mail	call	radio + web	mail	radio
(% who knew how)	3.12%	3.16%	3.16%	3.09%	3.14%
(standard deviation)	0.47%	0.44%	0.48%	0.36%	0.36%
Least effective	call	mail+radio+web	mail+radio+web	mail + web	all but ad
(% who knew how)	3.29%	3.35%	3.34%	3.23%	3.25%
(standard deviation)	0.39%	0.31%	0.38%	0.43%	0.38%

the 30-40 range were more likely to communicate among each other, but were equally likely to communicate with agents in the 40-50 range as in the 50-60 range. While age did not have a direct effect on the likelihood that an agent would want to communicate with the intervention – such decisions were entirely based on knowledge, belief, income level, and education level – the effects of literacy and information access differed greatly by age, as discussed in the technical report "Modeling Information Access in Construct" [1]. Thus, age had an indirect, though important, effect on agent knowledge and behavior.

The classes of agents most affected by age were the young (0-29) and the old (60+). Older agents were less likely to be literate: while an average of 88% of those under sixty-five were literate, only 77% of those who were older were able to read. Thus, older agents were more likely to suffer from the truncation effect of the intervention and would be less likely to receive knowledge or belief information from the advertisement, web site, or mailing. The trends with internet access were even more striking. While ninety percent of those under thirty went online, less than a third of those over sixty-five did so. This meant that the vast majority of older agents were not able to access the web site intervention directly, and were less likely to be affected by its stronger message. On the other hand, newspaper readership was highest among the older members of society: 58% of those over sixty-five read a print newspaper regularly as compared to only 32% of those under thirty. Thus, older agents – specifically, older and literate agents – would be the ones most affected by the advertisement.

6.1 Knowledge

When the experiment results were broken down into the five age groups examined in the virtual experiment, an interaction was observed between intervention types and age levels. This interaction was significant at the .10 level ($p < .085$), though the main effects of age and income were not found to be statistically significant. In most of the experimental cases examined, between 3.0% and 3.3% of the agents in the society had sufficient knowledge of the behavior to be able to perform it successfully. When interventions were modeled, fewer agents knew of the activity because the interventions often were able to deter agents from communicating with the promoter or passing along secondary information. In the most realistic case modeled, the no-intervention condition was often likely to have the highest (or among the highest) rates of knowledge in the society. Nevertheless, a reasonable number of interventions managed to lead to a decrease in the amount of knowledge known in the society, as plotted in Figure 2. While

Table 9: Greatest Changes in Belief, by Age

Case	agent age range				
	0-29	**30-40**	**40-50**	**50-60**	**60+**
Fewest holding belief	all	all	all	all	all but ad
(% who held belief)	3.07%	2.96%	2.97%	3.02%	3.06%
(standard deviation)	0.31%	0.42%	0.37%	0.36%	0.43
Second	all but ad	all but ad	all but ad	all but ad	all but ad
(% who held belief)	3.12%	3.00%	3.07%	3.12%	3.07%
(standard deviation)	0.38%	0.37%	0.35%	0.51%	0.48%
Third	radio + web	radio + web	radio + web	radio + web	radio + web
(% who held belief)	3.12%	3.12%	3.18%	3.14%	3.13%
(standard deviation)	0.34%	0.40%	0.43%	0.40%	0.37%
Least effective	call	radio	mail + web	ad	ad
(% who held belief)	3.60%	3.61%	3.55%	3.63%	3.55%
(standard deviation)	0.45%	0.42%	0.51%	0.41%	0.40%

Table 8 displays the three most effective interventions by raw score, Figure 2 plots the relative change from the no-intervention case for each of the ages modeled. Since each of the ages had a slightly different baseline rate of knowledge, interventions might be effective or ineffective for particular age brackets.

Several interventions or intervention combinations lead to decreases in the amount of knowledge in the society. For agents in the youngest group, those agents aged 0-29, the most effective intervention was the ad+web intervention combination, a combination which lead to a 4% decrease relative to the no-intervention case; this was followed by the all but ad intervention which lead to a slightly smaller improvement. For agents in their thirties, the lowest level was observed when the mailing and web site were active, closely followed by the case in which the radio and web were active. Knowledge was lowest for agents in their forties for the case where all interventions except the web site were modeled, though many agents in this age group lead to increases in knowledge relative to the no-intervention case due to the fact that the no-intervention baseline for this group was smaller than that for most of the other groups. For agents in their fifties, the advertisement lead to a substantial decrease in knowledge, though the all-intervention combination, and the mail intervention, and the web site intervention also lead to knowledge decreases as well. Among the eldest agents, the all-intervention combination was effective, though the vast majority of interventions and intervention combinations were successful as well.

The results suggest substantial variation in the effectiveness of different interventions on knowledge, though most interventions lead to knowledge decreases. The most consistent decrease came from the radio+web combination, which lead to decreases in three of the five age groups – those under 40 and those over 60 – and only small rises in the others. The all-intervention combination could also be considered to be among the most effective as well: it lead to larger decreases for some age groups – notably, those in the 30-40 age range and those older than 50 – but lead to larger increases in the other groups. The intervention that led to the largest increase in knowledge was the mail+radio+web intervention, though this was largely due to large knowledge increases in the 30-60 age rage.

6.2 Belief

While the interventions often did not have a large impact on knowledge, they did have a statistically significant impact on belief. The main effect for intervention type was highly

Table 10: Greatest Changes in Behavior, by Age

Case	agent age range				
	0-29	30-40	40-50	50-60	60+
Most effective	radio + web	all together	all but ad	all together	all together
(% who took action)	0.526%	0.523%	0.516%	0.520%	0.471%
(standard deviation)	0.142%	0.153%	0.161%	0.129%	0.116%
Second	ad + web	radio + web	radio + web	web site	radio + web
(% who took action)	0.540%	0.532%	0.531%	0.549%	0.518%
(standard deviation)	0.151%	0.173%	0.137%	0.148%	0.160%
Third	mail+radio+web	web alone	mail + web	ad alone	web alone
(% who took action)	0.541%	0.535%	0.535%	0.555%	0.549%
(standard deviation)	0.141%	0.161%	0.155%	0.131%	0.151%
Least effective	mail alone	radio alone	mail+radio+web	call center	mail alone
(% who took action)	0.630%	0.620%	0.632%	0.616%	0.641%
(standard deviation)	0.148%	0.161%	0.169%	0.186%	0.167%

statistically significant ($p < .0005$), though statistical support for a main effect for age or an interaction between age and intervention type were not found. This finding has been seen in earlier work, in which intervention combinations tend to strongly affect beliefs [2].

While the fluctuations by age group were not statistically significant, they are presented in tabular form in Table 9 and as relative percentages to the baseline level of belief in Figure 3 on page 38. For agents in the youngest age cohort, the all intervention combination and the all but ad lead to the greatest decreases in the number of agents who held the belief; the call center and mail interventions were the only two interventions that were ineffective. Though the all and all-but-ad intervention combinations were effective for agents of this age group, these combinations were less effective than they were for agents of older age cohorts. For the agents between age 30 and 39, the all and the all but ad interventions were the most effective, while the mail intervention and radio interventions lead to virtually no change. Agents between age 40 and 49 were also most affected by the all and all-but-ad combinations; this time, though, the ad and mail+web interventions were the worst performers. The all and all-but-ad were again effective for agents in the 50-60 age range; for this group, the ad intervention was the only one that led to a slight increase in belief. Older agents were also strongly affected by the all-but ad combo and the all-intervention combo; however, the effect was weaker for this group than it was on some others due to the decrease caused by literacy and the effect of Internet access. The radio and ad interventions had small undesired effects.

From the data in Figure 3, several trends are apparent. Intervention combinations are generally more effective at reducing belief than are solo interventions, a result that has been seen in previous work. A related result, though, is the fact that interventions which affect belief in one age cohort are likely to be effective for other age cohorts. For instance, while the all-intervention combination does not have quite as strong an effect on the youngest and the oldest agents, it tends to have a stronger effect than any of the alternative interventions or intervention combinations. Belief is affected by the information and access mechanisms, as seen by the fact that agents in the middle age cohorts (those least likely to be affected by the mechanisms) are most strongly affected by the interventions. Generally speaking, the ad, call center, radio, and mail interventions alone have relatively weak effects on belief, though when combined in various bundles with the web intervention they become substantially more effective. For all age groups, the all-intervention combination is the most effective, while the advertisement alone is generally least effective.

6.3 Behavior

Important differences in behavior were found for different interventions. The interaction effect observed for intervention and age on behavior was highly significant, ($p<.006$), suggesting that important differences were seen in terms of differing effectiveness of the interventions in deterring behavior. A main effect of intervention type was statistically significant ($p < .0005$), though a main effect for age was not found. The three most effective interventions are presented in tabular form in Table 10, which presents the percentage of the society which performed the action at the end of the simulated time period, while Figure 4 on page 39, marks the percentage change in behavior between each of the eleven intervention cases and the no-intervention case.

As can be seen from the data, the most effective intervention for agents in the youngest age was the radio+web intervention combination, which lead to a five percent decrease relative to the no-intervention case. Two other interventions also lead to a decrease: the ad+web combination and the mail+radio+web combination. The other interventions, including the all-intervention combination, lead to an increase in the number of agents who chose to perform the action; the least effective of these was the mail intervention, which led to a 14% increase in the number of agents who chose to perform the action. For the agents aged 30-39, the all-intervention combo was particularly effective, leading to a greater than 10% drop relative to the no-intervention case, but the radio+web combo and web intervention alone were also effective. For this age group, the vast majority of the interventions lead to a decrease in belief, although the radio intervention was one of four interventions which lead to about a 6% increase in the number of agents who performed the action.

Agents in their forties were most strongly affected in the case where all interventions but the advertisement were present – the advertisement lead to a much higher rate of activity for this age group, a surprising finding – with the mail+web combo and radio+web combo being more effective than others. These interventions lead to 4%-7% drops in the number of agents who performed the behavior. The mail+radio+web intervention, however, was particularly ineffective for this age group, as it lead more than 12% of the population to perform the action. Agents in their fifties were most affected by the all intervention bundle, which contributed to a 10% decrease in the number of agents who performed the action. Other bundles did not affect behavior as strongly: the next most effective interventions, the web site and the advertisement, lead to 5% and 4% decreases in behavior, respectively. Only the mail+web and the call center interventions lead to increases in behavior, however; most interventions lead to slight decreases in the number of agents who chose to perform the activity. The agents over sixty were most strongly influenced by the combined intervention bundle, which lead to a 20% drop in the amount of activity performed by these agents. The next most effective interventions, the radio+web and web site intervention, lead to decreases of about 8%. The mailing intervention was ineffective due to the illiteracy rate, which lead to a decrease in the number of agents who were able to absorb the entire message effectively.

While the most effective intervention varied slightly from age group to age group, several trends are apparent. First, the interventions that were generally effective included the all-intervention combination, the radio+web, and the web interventions. The all-intervention combination lead to decreases in activity for every group except for those in the 0-29 range; the decrease in behavior for the older agents was particularly striking as well as statistically significant. The radio+web intervention lead to a relatively consistent decrease in behavior, strongest for those in the 30-39 and 60+ age groups and weakest for those in the 50-59 group.

Table 11: Most Effective Interventions for Real Cities, by Age

Case	agent age range				
	0-29	30-40	40-50	50-60	60+
Stylized city	radio + web	all together	all but ad	all together	all together
(% of city pop)	20%	20%	20%	20%	20%
Hartford	mail + web	mail	web alone	all together	all together
(% of city pop)	38%	15%	15%	12%	19%
San Diego	mail + web	all together	mail	all but ad	mail + web
(% of city pop)	44%	18%	15%	10%	14%
Kansas City	all but ad	call center	call center	radio + web	web alone
(% of city pop)	43%	16%	10%	10%	16%
Orlando	ad alone	all but ad	radio + web	all but ad	all but ad
(% of city pop)	42%	17%	16%	11%	14%

The web site was generally not as effective as the all-intervention combination, though it lead to a similar pattern in terms of agent behavior. On the flip side, several interventions were seen to be relatively ineffective for all age groups. The mail intervention was generally ineffective, and the call center, advertisement, and radio interventions were washes as they were effective for some socio-demographics but not for several others.

6.4 Comparison to City Data

As a comparison case, the data obtained in this virtual experiment was compared to results obtained when simulating several real cities. The cities simulated – Hartford CT, San Diego CA, Kansas City, MO, and Orlando FL – were also simulated cities of three thousand agents, though the populations of the cities was built using data extracted from the PUMS data of the United States census. Unlike the stylized cities examined in the earlier portion of this experiment, this data included correlations between age, income, education level, number of children, and other socio-demographic attributes. Additional details about the cities simulated can be found in the technical report "The Impact of Educational Interventions on Real and Stylized Cities" [2].

Table 11 presents the twenty-five most effective interventions, the single most effective intervention for the stylized city and each of the four real cities, and each of the five age brackets. Comparing the first row of Table 11 to the remaining rows, several trends are apparent. First, there are substantial differences in the fraction of the populations in each age bracket. While the stylized city had five brackets of roughly equal proportions, over four in ten agents were in the youngest age bracket. Second, there were often a substantial difference between the interventions which appear to be effective for the stylized city and each of the real cities. For example, the interventions which appeared to be most effective for the youngest agents in the real cities were not in the top three for the stylized city. Nevertheless, there were some overlaps in several socio-demographic groups: for example, the results for the older agents in Hartford and Kansas City were similar to those observed in the stylized cities.

It is important to note that there are several differences between the results obtained in Table 11 and in the preceding experiment. First, due to the correlations between different socio-demographic parameters, agents in the real cities were less likely to interact with agents of differing socio-demographics than in the stylized cities. Since socio-demographics were often correlated, agents would have greater in-group affinity and would be less likely to interact with agents of a different socio-demographic group. In the stylized city with uncorrelated parameters, agents were as likely to differ on one parameter as any other, leading to less pronounced socio-demographic differences. Thus, some of the differences between Table 11 and the stylized results are due to different fundamental interaction patterns. Additionally, though, it is important to note that only the single most effective intervention for each city and each age group is presented. When examining the second- and third-most effective interventions for each of these categories, substantially more overlap is observed. Due to the relatively large impact of individual fluctuations – one single interaction could snowball to ten agents performing the action in one replication due to the spreading of information – Construct is not necessarily designed to identify the single best available intervention bundle in any situation, but rather can identify a suite of interventions or intervention bundles that would be likely to perform better than others in most situations.

7 Specific Effects by Income

The second socio-demographic variable examined was income, the measurement of how much a simulated agent earned. Six general income brackets were used: annual incomes in the $0-15k range, $15-30k range, $30k-50k range, $50k-80k range, $80k-120k range, and greater than $120k range. The bottom two ranges were selected for their initial correspondence to the generalized cutoffs used for the Earned Income Tax Credit, while the remaining bin sizes were decided in conjunction with simulation experts in order to represent a relatively broad set of agents and households [7]. Unlike age, discussed in Section 6, income had a direct effect on how likely agents would be to communicate with an intervention: agents in the lowest two income levels had a predisposition to communicate since the interventions were geared for agents at that level. However, income also had an important effect in terms of information access. Income has been shown to be strongly correlated with literacy and information access, as discussed in the technical report “Modeling Information Access in Construct” [1]. While low income individuals were more likely to interact with the intervention, however, they were less likely to be able to do so.

It has been observed that high income individuals have a much higher rate of literacy, internet access, and newspaper readership than low-income individuals. In the Construct simulation, this was modeled by allowing a greater percentage of high-income individuals to interact with the modeled interventions. The literacy rate for those making over $75,000 a year in the United States was well over 99%, as compared to 82% for those making less than $30,000 [8]. This meant that almost all upper-income agents were literate and did not suffer from the truncation and misinformation effects described in Section 3.2, while one in five of the low-income agents did. Since over nine in ten Americans with income over $75,000 go online “at least occasionally,” as compared with less than six in ten of those whose income was less than $30,000, these parameters were used for Construct agents with similar income levels [9]. Lastly, the newspaper readership rate was substantially greater – 58% versus 40% – for the two income brackets, a factor that was also represented in the simulation [10]. These numbers suggested that low-income agents would be least likely to have direct access to, and thus be less affected by, interventions which required any of the literacy or access constraints.

It is important to note that, as described in Section 4.1, only a subset of the agent population was used in the calculation of the statistics. Since the action being simulated was designed primarily for low-income individuals, individuals with the lowest level of income were most likely to seek out the promoter or the intervention. Only a relatively small percentage of high-income agents were able to “match” the behavior, meaning that they could potentially perform the action; non-matching agents were included in the society to serve as potential conduits of information to both high- and low-income agents. The percentages reported in Sections 6 to 8 reflect only these matching agents, except where otherwise noted.

Before presenting the results, it is necessary to observe that the differences in population – a large number of low income agents with only a small number of high-income individuals – lead to uneven variances in the results. These unequal variances were not helped by the treatments, since the effects of literacy and information access were not always the same in each subpopulation. As a percentage of the mean, then, the standard deviations are quite small for low income levels, but can increase tenfold for high-income agents. Thus, it is necessary to view the mean scores as less reliable for the high income agents, as well as to recognize that the differences in variance may not be directly comparable since the fractions used were different.

7.1 Knowledge

Table 12: Greatest Changes in Knowledge, by Income

Case	**agent income level**					
	$0-15k	**$15k-30k**	**$30k-50k**	**$50k-80k**	**$80k-120k**	**$120k+**
Fewest knowing how	ad	web	mail	mail	mail	all
(% who knew how)	3.16%	3.08%	2.86%	2.74%	3.27%	2.42%
(standard deviation)	0.26%	0.28%	0.67%	0.74%	1.32%	1.80%
Second	call	all	ad + web	all	mail + web	all but ad
(% who knew how)	3.17%	3.10%	2.96%	3.01%	3.12%	2.74%
(standard deviation)	0.31%	0.27%	0.54%	0.75%	1.01%	2.63%
Third	radio + web	mail + web	radio + web	web	all but ad	mail
(% who knew how)	3.18%	3.13%	3.04%	3.12%	3.14%	2.98%
(standard deviation)	0.36%	0.30%	0.56%	0.97%	1.34%	2.19%
Least effective	mail+radio+web	mail	mail + web	none	ad + web	mail+radio+web
(% who knew how)	3.27%	3.25	3.34%	3.49%	3.50%	3.87%
(standard deviation)	0.33%	0.39%	0.72%	1.06%	1.37%	1.97%

Due to the differences observed in the standard deviations around the means, as can be seen in the six different columns of Table 12 a weighted least squares procedure was used to correct for unequal variance. A Levine's test had demonstrated that the sample variances were problematic for a standard ANOVA analysis ($p < .0005$), an effect that was likely due to the smaller population sizes used for high income individuals. With these variance corrections, statistically significant main effects were found for both income ($p < .039$) and intervention type ($p < .0001$). A statistically significant interaction term was noted as well ($p < .0005$). This suggested that important differences were observed both for different income levels overall, but more importantly that different interventions could lead to different effects for each income level.

Table 12 displays the three most effective interventions for decreasing knowledge as well as the single least effective intervention for decreasing knowledge in the society. For agents in the lowest income bracket, the advertisement was most effective, followed by the call center, followed by the radio+web combo. However, the difference between the most effective and least effective interventions was not very large, as can be seen in Figure 5 on page 40. The advertisement, for instance, which lead to 3.16% of the society being informed represented only a 1% decrease from the no-intervention case; only three interventions or intervention combinations contributed to slight decreases in knowledge. Agents in the $15-30k range were most affected by the web intervention, the combined intervention, and the mail+web intervention. These interventions managed to decrease knowledge about 3% relative to the no-intervention case, a stronger effect than that observed for the lowest income level. While the cognitive and access constraints may have limited the exposure of these agents to the knowledge contained in these interventions, this effect was not pronounced.

Agents in the higher income brackets were neither targeted by the promoter nor were targets of the intervention; they also did not necessarily meet the matching criteria and many were ignored. However, middle- and high-income agents were still able to learn a substantial amount of information about the scheme, and the interventions could inform or modify their behaviors as well, and – as Figure 5 illustrates – could lead to large changes relative to the no-intervention case. Agents in the $30k-50k range, were most affected by the mail, ad+web, and radio+web interventions, while they were least affected by the mail+web intervention combination. Agents in the $50k-80k range were most affected by the mail intervention, the all intervention combination, then the web intervention; for this demographic group, all interventions were effective at decreasing knowledge. For the higher income levels, those making more than $80k, the mail intervention and the all-but-ad combination were generally effective interventions; the ad+web and the mail+radio+web combinations were particularly ineffective.

Several trends are apparent in the data. On the one hand, there is substantial variation between the different types of interventions by income level. For instance, of the three best interventions for the lowest three income levels, eight different intervention combinations are effective. This suggests that there is not necessarily a correlation in terms of the most effective intervention by income level for these groups. This many be due to the uneven effects of the information access constraints, some of which can affect interventions and intervention combination sand prevent low-income agents from learning information about the activity. On the other hand, the mail intervention does decrease knowledge for the upper-income groups, as it is the most effective for the agents in the $30k-120k range and the second-best for those in the highest range. Similarly, the all-intervention combination can be effective for several of the income groups. This suggests that high-income agents are less likely to be unevenly affected by the information access mechanisms, which may allow certain interventions to be more consistently effective.

7.2 Belief

Table 13: Greatest Changes in Belief, by Income

Case	**agent income level**					
	$0-15k	**$15k-30k**	**$30k-50k**	**$50k-80k**	**$80k-120k**	**$120k+**
Fewest holding belief	all	all but ad	web	mail+radio+web	all	mail+radio+web
(% who held belief)	3.05%	2.99%	3.11%	2.78%	2.58%	2.63%
(standard deviation)	0.29%	0.32%	0.55%	0.74%	1.20%	1.98%
Second	radio + web	all	all	all	web	all
(% who held belief)	3.15%	3.00%	3.12%	2.87%	3.04%	2.80%
(standard deviation)	0.29%	0.27%	0.43%	0.80%	1.13%	2.20%
Third	all but ad	radio + web	all but ad	radio + web	all but ad	all but ad
(% who held belief)	3.15%	3.13%	3.13%	2.92%	3.08%	3.09%
(standard deviation)	0.30%	0.27%	0.61%	0.70%	1.42%	2.13%
Least effective	call	mail	none	none	ad	ad
(% who held belief)	3.56%	3.56%	3.48%	3.54%	4.43%	4.20%
(standard deviation)	0.32%	0.32%	0.55%	0.93%	1.42%	2.42%

For belief, a statistically significant main effect was observed for intervention type ($p < .0005$). However, a statistically significant main effect was also observed for income level ($p < .014$), with important decreases noted in the middle and upper income levels. The interaction between belief and income level was highly significant as well ($p < .005$), however, suggesting important differences between the intervention types, the income levels, and the mean level of belief in the society. As before, the least squares regression was used in order to account for the increased variance observed as income levels increased.

Table 13, as well as Figure 6 on page 41, displays the three most effective interventions on belief when broken down by income level. Individuals at the lowest income levels were most affected by the all-intervention combination, then the radio+web intervention, then the all-but-ad intervention. All three of these interventions decreased belief by about 10% relative to the no-intervention condition. The call center, however, contributed to a slight increase in belief. Individuals in the second-lowest income level were affected by the same interventions but in a different order: the all-but-ad combination was the most effective, followed by the all-intervention combination, followed by the radio+web. These interventions lead to a 15%-20% decrease in belief relative to the no-intervention, suggesting a much stronger overall effect. In this case, the mail was the least effective intervention, contributing to a very slight rise in the total number of agents who held the belief.

Higher income agents were also affected by the interventions, but were affected slightly differently. The all intervention and the all-but-ad intervention appeared to be effective for almost every income bracket, though the ordering of the interventions varied by income level. The web intervention was effective for those in the $30k-$50k range, as well as those in the $80k-120k; the mail+radio+web intervention was effective for the $50k-$80k and $120k+ agents. Most interventions were generally effective, often leading to 20% decreases in belief, though the advertisement was a notable exception. As before, though, the variance increased for agents with higher incomes, meaning that the mean values were less reliable for the high-income agents as compared to those with lower incomes.

As can be seen from Table 13, the all-intervention combination was the most effective combination for reducing belief. The all-but-ad intervention was also effective, though in most cases it was less effective than the all-intervention combination. The web, radio+web and mail+radio+web interventions also tended to reduce belief substantially, though their effects depended more on the type of intervention used. This suggested that, for belief, the combinations were still likely to be the most effective interventions as they could convey information to agents who did not suffer from the access constraints. Since these agents could then relay information to other agents in the society – including those agents who were illiterate, who lacked internet access, or who lacked newspaper access – the interventions could still affect agents in these socio-demographics.

7.3 Behavior

An interaction between income level and intervention effectiveness was observed for behavior ($p < .0005$), as it was for knowledge and behavior. This suggested that the different interventions could affect different societies in different ways. A main effect of income level was not found, though there was a statistically significant difference by intervention ($p < 0.25$). As can be seen in Table 14, there were slight changes in behavior due to the impact of different interventions on different sub-populations. It is important to note an important overarching trend as illustrated: as income decreases, the percentage of agents who perform the behavior increases even for the most effective interventions. While 0.151% of the matching agent population performed the action in the case of the most effective intervention for high-income agents, 0.542% of the population did so for the low income agents. Low income agents were most likely to be illiterate, most likely to lack internet access, and most likely not to have a newspaper subscription; in short, they were the ones who were least likely to be directly affected by the educational interventions overall. However, the low-income agents were the ones most likely to be contacted by the promoter, since they were the target demographic for which the behavior was designed. Therefore, while some of the increase can be seen as a failure due to literacy and restricted information access, some of the difference is due to specific targeting by the promoter as well.

Table 14: Greatest Changes in Behavior, by Income

Case	**agent income level**					
	$0-15k	**$15k-30k**	**$30k-50k**	**$50k-80k**	**$80k-120k**	**$120k+**
Most effective	ad	all	radio + web	all	mail	mail
(% who took action)	0.542%	0.509%	0.503%	0.403%	0.377%	0.151%
(standard deviation)	0.090%	0.091%	0.267%	0.283%	0.441%	0.547%
Second	all	radio + web	all	mail	mail + web	radio
(% who took action)	0.548%	0.519%	0.508%	0.417%	0.468%	0.353%
(standard deviation)	0.098%	0.105%	0.164%	0.330%	0.496%	0.784%
Third	radio + web	web	web	all but ad	radio	ad + web
(% who took action)	0.556%	0.535%	0.535%	0.479%	0.511%	0.491%
(standard deviation)	0.123%	0.100%	0.238%	0.356%	0.422%	0.795%
Least effective	call	mail	ad + web	radio	all	mail+radio+web
(% who took action)	0.604%	0.651%	0.626%	0.573%	0.695%	0.714%
(standard deviation)	0.170%	0.127%	0.225%	0.476%	0.635%	1.014%

The relative effectiveness of the interventions for different agent groups is also interesting. For agents in the lowest income bracket, as illustrated numerically in Table 14 and graphically in and Figure 7 on page 42, the advertisement was the most effective intervention as it decreased the behavior to 0.542% of the population. This represented a 3.5% decrease in the number of agents who performed the action relative to the no-intervention case. Much of this decrease was likely driven by a corresponding decrease in the amount of knowledge in the society. As illustrated in Table 12, the advertisement lead to a decrease in knowledge in the society, leaving fewer knowledgeable agents to potentially perform the behavior. The next most effective interventions, the all-intervention combination and the radio+web intervention combination, were largely driven by decreases in belief as opposed to knowledge. However, these were the only three interventions which lead to a mean decrease in the number of agents who performed the action; most of the rest of the interventions and intervention combinations lead to slight increases. The least effective intervention was the call center, which lead to an 8% increase relative to the no-intervention case.

Agents in the second-lowest income bracket, the $15k-30k range, were affected by a similar set of interventions. The all-intervention combination was particularly effective, leading to a 13% decrease in the number of agents performing the action relative to the no-intervention case. The radio+web combination also lead to double-digit decrease relative to the no-intervention case. These decreases, like the decreases for the lowest-income agents, were helped by the large decreases in belief observed in Table 13. Only two interventions did not lead to net decreases in the percentage of agents who performed the behavior relative to the no-intervention case, the advertisement and the mail intervention. The mail intervention lead to a 10% increase in the number of agents performing the behavior relative to the no-intervention case, a result that was likely due to the increase in knowledge observed for this intervention.

For agents at higher income levels, the effects on behavior were mixed. For agents in the $30k-50k range, the radio+web combination and all-intervention combinations were effective, leading to decreases of over ten percent relative to the no-intervention case, while the web site was the most effective single intervention leading to a decrease of about eight percent relative to the no-intervention case. The ad+web combination was ineffective and lead to a 12% increase in activity. For the agents in the $50k-80k range, the all-intervention combination and mail intervention decreased behavior by about 20%, while the all-but-ad intervention decreased activity by about 10%. The radio intervention was one of three interventions which lead to a very large increase in behavior, an increase of about 25%. For agents in the highest income brackets, the mail interventions and radio interventions lead to the greatest decreases in behavior, while the all combination and mail+radio+web combinations were ineffectual.

It is important to note that the standard deviations for high income levels were extremely high relative to the mean. For this reason, the mean values for behavior were not always useful as predictive tools. Small fluctuations in the underlying social network could have a large impact on the number of high-income individuals who perform the action, especially if few of the agents existed in the first place. For this reason, the data reported for the high-income agents should be treated with extreme caution. The standard deviations for low-income agents are much smaller relative to the mean, and therefore the means for the $0-15k, $15-30k, and even the $30-50k are more reasonable, and the mean outcomes for these socio-demographic groups are reasonable representations of likely outcomes.

7.4 Comparison to City Data

The breakdown by income level for the stylized city was compared to four real cities, the same cities used for comparison in Section 6.4. This data is presented in Table 15, with the most effective intervention in Hartford, San Diego, Kansas City, and Orlando paired with the percentage of the city's population that is within that income range. It is important to note that the most effective interventions computed for Table 15 were computed for the entire society, and not for a subset of matching agents. This meant that additional high-income agents were considered when determining the most effective interventions for high-income agents and served to slightly decrease the variance.

Table 15: Most Effective Interventions for Real Cities, by Income

Case	agent income level					
	$0-15k	$15k-30k	$30k-50k	$50k-80k	$80k-120k	$120k+
Stylized	ad	all	radio + web	all	mail	mail
(% of city pop)	40%	35%	15%	6%	3%	1%
Hartford	mail + web	radio	web	all but ad	mail+radio+web	mail+radio+web
(% of city pop)	53%	16%	16%	10%	3%	3%
San Diego	mail + web	all	all	mail + web	radio + web	call
(% of city pop)	56%	17%	12%	7%	3%	3%
Kansas City	web	mail	call	radio + web	all but ad	call
(% of city pop)	56%	20%	15%	6%	1%	1%
Orlando	ad	all but ad	radio + web	mail + web	web	mail
(% of city pop)	56%	20%	13%	6%	2%	2%

Comparing the stylized city, the first row in Table 15, with the real cities in the remaining rows, more differences are apparent than similarities. However, this was not necessarily the case. While the advertisement was most effective in the stylized city and in Hartford for the low-income agents, and while the all-intervention combination was the most effective in the stylized city and in San Diego for the $15k-30k range, this finding did not generalize to the other cities. For the $0-$15k population, Hartford and San Diego both found the mail+web intervention to be effective, while it was only a mid-range intervention for the stylized city. The web site, the most effective for Kansas City, was in the top half for the stylized city. The results were similar for agents in the $15k-30k population. While the mail intervention was the least effective intervention for the stylized city, the all-but-ad intervention was the fourth-most effective intervention for the stylized city and the radio was in the upper half. These relationships, however, are not captured in the table because only the top interventions are captured.

Three general themes can be seen, however. First, the same intervention was not necessarily the most effective for all income levels. While the all-intervention bundle and the mail intervention tended to be the most effective interventions for the stylized city, only San Diego had a general set of effective interventions – the mail+web and the all-intervention combo. Other cities tended to have different interventions that were more effective for sub-populations but not for others: Orlando, in particular, had a different effective intervention for each of the six income levels. When the second- and third- most effective interventions were considered, however, there was slightly more consensus within and between cities. The web site alone showed up as one of the most effective interventions for agents making less than $50k in most cities, primarily due to its large messages and the effect it had on decreasing belief, while the advertisement and the all-intervention combinations were generally effective for those making under $30k a year.

A second major theme was that intervention bundles were not always successful. While all of the most effective interventions for the $50k-80k agents were intervention bundles, for the $0-15k agents this was true only in Hartford and San Diego. For low income agents, more interventions did not necessarily lead to a decrease in behavior: the ad, web, mail, and radio interventions were each effective for low-income populations in different cities. This contrasts with the societal slices by age – as discussed in Section 6 – which suggested that most of the effective interventions were bundles. This seeming contradiction can resolved by the examining the binning process more closely, since income and age were correlated but not identical. Indeed, many of the intervention combinations present in Table 11 were among the upper half of the interventions when the data was sliced by income, though because they were not in the single most effective interventions.

Lastly, when only the most effective bundles were considered, the smaller bundles often outperformed the larger bundles. For instance, the radio+web and the mail+web bundles were almost as effective as the all-intervention bundles, especially in certain cities and for certain demographics. While the all-intervention bundle and the all-but-ad bundle were often the second- or third-most effective interventions in many situations, the smaller bundles tended to come out ahead.

8 Specific Effects by Education

Other researchers have show that education is strongly correlated with literacy and information access, as discussed in the technical report “Modeling Information Access in Construct” [1]. Four education levels were examined in this virtual experiment – less than high school education, high school education or high school diploma, some college or college degree, and some graduate / professional training or graduate / professional degree. These bins were chosen largely due to the way other social science data, such as the literacy rates and information access rates, were recorded, although these four bins had also been used in past virtual experiment work with Construct [2][8]. Like income, discussed in Section 7, education affected both an agent's ability to interact with and gain knowledge from an educational intervention, and its propensity to do so. Since the behavior modeled was designed for low-income individuals and those who did not have a graduate degree, agents in all but the highest education level had an increased propensity to communicate with the promoter and educational interventions. However, since low-education agents were least likely to be literate or have access to newspapers on the internet, agents with lower levels of education were not always able to be directly affected by the educational interventions.

Education is one of the strongest predictors of literacy rates. Only about 50% of Americans with less than a high school education were considered to be above a "below basic" level of literacy as of 2003; in contrast, literacy was nearly universal for those who had post-collegiate education [8]. Similarly, while only 40% of Americans with less than a high school education used the internet occasionally, nearly nine in ten highly-educated Americans did [9]. Newspaper readership was also strongly correlated with education: while less than three in ten Americans with less than a high school education read an issue of a newspaper a week, nearly twice as many highly-educated Americans did so [10]. These numbers were used in the simulation to reflect how education played an important role in determining whether or not individuals learned information from different types of educational interventions.

It is important to note that, as described in Section 5, only a subset of the agent population was used in the calculation of the statistics. Since the action being simulated was designed primarily for low-education individuals, individuals with less than a college education were more likely to seek out the promoter or the intervention. Only a relatively small percentage of agents in the upper education brackets were able to "match" the behavior, meaning that they could potentially perform the action; non-matching agents were included in the society to serve as potential conduits of information to both high- and low-income agents. The percentages reported in 7.1 to 7.3 reflect only these matching agents, a slightly greater percentage than that taken from the whole population.

8.1 Knowledge

Table 16: Greatest Changes in Knowledge, by Education

Case	**agent education level**			
	less than hs	**hs diploma**	**college degree**	**grad school**
Fewest knowing how	radio + web	all	call	ad + web
(% who knew how)	3.13%	3.10%	3.13%	2.61%
(standard deviation)	0.32%	0.33%	0.52%	2.10%
Second	mail	mail + web	ad	mail
(% who knew how)	3.14%	3.11%	3.14%	2.77%
(standard deviation)	0.32%	0.35%	0.47%	2.09%
Third	all	web	mail + web	mail + web
(% who knew how)	3.14%	3.15%	3.15%	2.81%
(standard deviation)	0.25%	0.34%	0.48%	2.18%
Least effective	mail + web	mail + radio + web	all	radio
(% who knew how)	3.24%	3.31%	3.27%	3.63%
(standard deviation)	0.33%	0.35%	0.40%	2.38%

When examining breakdowns by education, there were large differences in the variance. Because the percentage of the population that had a graduate education was much smaller than the other education groups, the variance observed there was greater since minor fluctuations to this group produced larger overall changes. However, once corrected for variance, the only statistically significant difference observed was that for intervention, and even that was only significant at the .10 level ($p < .081$). The main effect for education, and for the interaction between education and intervention type, were not statistically significant due to the fact that the differences between the observed means was small relative to the variance endogenous to the Construct simulation. As can be seen in Table 16, variance stemming from the underlying social networks, the initial fact assignments, and other random factors was substantially greater than any of the differences between each education level.

Table 16 and Figure 8 (on page 43) demonstrate that different interventions had different effects on knowledge. Though not statistically significant, these trends are presented for consistency and to motivate further study. For agents with less than a high school education, the radio+web intervention was the most effective, decreasing knowledge by about 3% relative to the no-intervention case. The mail intervention and the all-intervention combination were also effective, while the mail+web combo was not. For agents with at least a high-school diploma, the all-intervention combination was effective, also decreasing knowledge 3% relative to the no intervention case. For these agents, however, the mail+web intervention combination and the web intervention were effective, while the mail+radio+web intervention combination lead to a net increase in knowledge relative to the no-intervention case. The call center, advertisement, and mail+web interventions lead to a 1% decrease in the amount of knowledge in the society for agents with at least a college degree, while the all-intervention combination lead to slightly less than a 3% increase. Agents who had advanced educations were strongly affected by the ad+web, mail, and mail+web interventions, which lead to a 20% decrease relative to the no-intervention case. However, because the population for these agents was small in the society, the variance around this mean was much larger than those for the other three education levels.

The lack of a substantial statistically significant effect for knowledge suggests that no major changes occurred due to knowledge. This may at first seem surprising, given that there were statistically significant interactions by age (Section 6.1) as well as statistically significant interactions and main effects by income (Section 7.1). However, this is likely due to the differences in data slicing, as well as the lack of correlation among agent socio-demographic parameters. While all three sets of statistics were computed from the underlying data source, each socio-demographic parameter represented a different slice of the same data. Each socio-demographic slice had a different composition, and since the data was analyzed as a percent of the socio-demographic group which was knowledgeable, the denominators were not necessarily equal between different socio-demographics. Additionally, though, the lack of correlation among socio-demographic attributes may also have contributed to this effect. Since there was no guaranteed that low education agents were likely to be low income, a well-documented correlation in the literate, there was no guarantee that socio-demographic groups were likely to have been affected by the same socio-demographic properties. Thus, even though slicing the results by income may have lead to one type of effect that was seen to be statistically significant, slicing by a different parameter might not.

8.2 Belief

After correcting for the differences in variance, the only factors which influenced belief were found to be education level ($p < .003$) and intervention ($p < .0005$). The interaction effect between the two parameters was significant at the .10 level ($p < .095$), but did not have the predictive power that it did for the other socio-demographic breakdowns such as age or income. and appendix Figure 9, shows how different interventions affected the beliefs of different groups.

Table 17: Greatest Changes in Belief, by Education

Case	agent education level			
	less than hs	**hs diploma**	**college degree**	**grad school**
Fewest holding belief	all	all	all	ad + web
(% who held belief)	3.00%	3.07%	2.93%	2.88%
(standard deviation)	0.22%	0.30%	0.38%	2.38%
Second	all but ad	all but ad	all but ad	radio + web
(% who held belief)	3.05%	3.11%	3.04%	3.29%
(standard deviation)	0.32%	0.35%	0.44%	2.30%
Third	radio + web	mail + radio + web	radio + web	web
(% who held belief)	3.06%	3.14%	3.26%	3.30%
(standard deviation)	0.27%	0.38%	0.47%	2.15%
Least effective	mail	call	ad	mail
(% who held belief)	3.58%	3.58%	3.53%	3.86%
(standard deviation)	0.34%	0.34%	0.48%	2.42%

For agents with less than a high school education, a high school diploma, or college degree – 99% of the simulated society – the most effective educational intervention was the all-intervention bundle. Relative to their respective no-intervention cases, the all-intervention combination lead to a 15% decrease in belief for those with less than a high school education, a 12% decrease in belief for those with a high school diploma, and 17% decrease for those who had a college degree. While the all-but-ad bundle was the next most effective bundle for all three agent groups, the remaining interventions affected each group differently. The agents with a graduate school education had a different pattern of most and least effective interventions. For these agents, the ad+web intervention led to a 25% decease in the number of agents who held the belief relative to the no-intervention case; the next-most effective was the radio+web followed by the web alone.

As can be seen from the table, the most powerful interventions for affecting beliefs were almost exclusively combinations. In most cases, the most effective interventions for reducing belief were the ones that had the largest amount of information. This occurred even for the low-income agents that were more likely to be affected by illiteracy and lack of information access. This suggests that communication between high-income agents and low-income agents was able to pass along a substantial amount of belief information to these agents, that the smaller number of low-education agents with full information access were able to relay information to their peers, and that more interventions were generally helpful for encouraging this information sharing.

8.3 Behavior

Table 18: Greatest Changes in Behavior, by Education

Case	agent education level			
	less than hs	**hs diploma**	**college degree**	**grad school**
Most effective	all	all	radio + web	radio + web
(% who took action)	0.512%	0.535%	0.525%	0.331%
(standard deviation)	0.083%	0.097%	0.169%	0.706%
Second	radio + web	radio + web	all but ad	web
(% who took action)	0.537%	0.545%	0.530%	0.382%
(standard deviation)	0.104%	0.129%	0.176%	0.733%
Third	web	web	ad	all
(% who took action)	0.549%	0.547%	0.552%	0.401%

(standard deviation)	0.084%	0.117%	0.186%	0.707%
Least effective	mail	mail	mail + web	all but ad
(% who took action)	0.611%	0.609%	0.601%	0.760%
(standard deviation)	0.115%	0.113%	0.169%	0.871%

When a weighted least squares regression was run for behavior, the main effect for intervention level was found to be weakly statistically significant at the .10 level ($p < .092$). The main effect for education level, as well as the interactions between education level and intervention, was not found to be statistically significant. Here again, the variances due to differences in social networks, initial knowledge, and other factors in the simulation lead to variances that were substantially greater than any of the effects of the cognitive or access changes. These findings are summarized as percentages of the population in Table 18 and as percentages relative to the no-intervention case in appendix Figure 10.

For individuals with less than a high school education, the all-intervention combination lead to the greatest decrease in behavior, a value that was 12% less than in the case when no intervention was present. The radio and the radio+web interventions were the next most effective, while the mail intervention was the worst of four interventions which lead to a net increase in behavior over the no-intervention case. A similar trend was found for agents who had a high school diploma. For these agents, the all-intervention combination contributed to a 7% decrease from the no-intervention case, while the radio+web combo and web intervention lead to slightly smaller decreases. Agents with a college degree were most affected by the radio+web intervention, which lead to a 6% decrease as compared to the no-intervention case; the all-but-ad and the ad interventions were found to be the next most effective for these sub-populations. For the population with the highest level of education, the radio+web intervention was also found to be the most effective, while the web intervention and all-intervention combination were found to be effective as well. Thus, while the results are not statistically significant, the experiments run provide a moderate amount of consensus that the all intervention, radio+web intervention, and the web intervention are generally effective at decreasing the amount of behavior in the stylized society.

8.4 Comparison to City Data

As was done for age and for income, the stylized city was compared to the same four real cities. The most effective interventions, as well as the percentage of the population which matched each criteria, is reported in Table 19.

Table 19: Most Effective Interventions for Real Cities, by Education

Case	**agent education level**			
	less than hs	**hs diploma**	**college degree**	**grad school**
Stylized city	all	all	radio + web	radio + web
(% of city pop)	40%	30%	29%	1%
Hartford	all but ad	radio	ad + web	all
(% of city pop)	39%	36%	17%	9%
San Diego	all	ad + web	radio	call
(% of city pop)	36%	36%	20%	8%
Kansas City	web	all but ad	call	mail
(% of city pop)	39%	40%	15%	5%
Orlando	call	radio + web	all but ad	ad
(% of city pop)	37%	37%	20%	6%

As can be seen, there was still a substantial difference between the stylized city and the real cities modeled. The greatest overlap occurred for agents with less than a high school education, where the San Diego and Kansas City populations were affected by one of the interventions noted as very effective in Table 18. For that demographic group, the all-intervention case was the most effective and the web intervention was the third-most effective. There were also overlaps with other education levels and other cities: Orlando's radio+web intervention was the second-most effective intervention for the stylized city, Orlando's all-but-ad intervention combination was the second-most effective intervention for those with college degrees, and Hartford's all-intervention combination was the third-most effective intervention for those with a graduate education.

As can be seen however, there were substantial variations between the population percentages in the stylized cities as compared to the real cities. While the stylized city had a greater percentage of agents with a college degree relative to the real cities, it had slightly smaller percentage of agents who had completed high school or who had some graduate training. This could have contributed to the differences observed between the real and stylized cities. Similarly, the correlation between the parameters in the real cities could also contribute to the differences in results.

9 General Observations

The experimental results discussed in Sections 6 through 8 have highlighted the fact that interventions may affect one subpopulation more than another. While some general trends can be drawn from the tables describing behavior – Table 10, Table 14, and Table 18 – most striking factor of these tables is that interventions that affect one sub-population may not be the most effective intervention for another sub-population. While the Construct simulation environment has a huge parameter space and only a small fraction of that space could be explored in this study, several of the results seen here suggested that there were statistically significant main effects or interactions for intervention type among these sub-populations. While the all-intervention combination and radio+web interventions was generally seen to be a successful intervention for reducing the behavior in these populations, the exact ordering of interventions could vary considerably from population to population. Along a similar line, substantial evidence suggested that the all-intervention combination and the all-but-ad combination were the most effective at discouraging belief largely due to the large amount of anti-behavior belief information that they possessed. The most effective interventions for reducing knowledge, however, was not clear: in some cases, intervention bundles were effective since they might provide enough belief information to discourage contact with the promoter; in other cases, weaker interventions like the advertisement intervention were effective since they provided as little additional information as possible.

The results of these sections also suggest that the most effective interventions for reducing any particular quality in the society were not necessarily the best methods for discouraging the behavior as a whole. While an intervention might decrease belief for a sub-population, it might not be effective at decreasing behavior. On the other hand, if it was effective at decreasing both belief and behavior, it might not be as effective for a slightly different socio-demographic group. For example, consider the effects of the all-but-ad intervention bundle on different age groups, as presented in Section 6. While it was universally seen as the second-best intervention for decreasing belief information in the society, it was not always the one that decreased the behavior effectively. For some age groups, such as those under thirty, it increased decreased knowledge due to less interaction with the promoter; in other age groups, such as those over sixty, it increased knowledge due to the information coming from the intervention cocktail. Nevertheless, the intervention was only effective for those agents in the 40-50 age range; for other agent populations, the intervention was not among the top three most effective interventions.

While some of the factors which lead to the differences in intervention communication were socio-demographic in nature – agents from low-income sub-populations were more likely to contact the intervention as compared to high income agents – many of the other differences between socio-demographic populations were due to changes stemming from information access parameters. While the effectiveness of various interventions changed for the different sub-populations, the magnitude of these effects was not very large. Most changes in knowledge and belief could be measured in fractions of percentage points, while the differences in behavior usually represented an extra agent in ten thousand. Thus, while the literacy and access parameters could have a large effect on an individual agent's knowledge, belief, or action, it had a relatively small effect on the aggregate society as a whole. At the level of analysis performed here, the population level, the changes seen were small.

However, within socio-demographic groups that were equally likely to communicate with the promoter, there was evidence that the cognitive and information access limitations had an important effect on the overall level of knowledge and belief in the simulated societies. For instance, consider the interaction between agent and intervention effectiveness on belief as modeled in Table 9. Of these sub-populations, the elderly were most affected by illiteracy and lack of internet access, while the younger agents were most affected by a lack of newspaper access. The all-intervention combination, the most effective intervention for all age groups, affected each sub-population differently. The agents in the middle age ranges, who were most likely to have literacy and information access, had the lowest overall belief rates at about 2.97%. The young and the elderly, who were least likely to have direct access to the intervention, were both a tenth of a percentage point higher. While this did not directly translate into fewer people performing the action, it suggested that the lack of direct access to the intervention could have an important effect on the aggregate percentage of the population who had a particular attribute of interest.

However, the differences in knowledge, belief, and communication were often more complex than can be explained by the effects of a sub-population's literacy and access percentages alone. Though agents of a sub-population may be less likely to learn information directly from an educational intervention, they may be able to obtain the information via communication indirectly via agents of other sub-populations. This could result in two important effects, both of which are difficult to distinguish from the simulation results.

First, information access parameters could fail to lead to a change in the affected population. While the socio-demographic characteristics of any agent who initially interacts with the intervention are important, the interactions of that agent are equally important. A lack of information access could prevent an agent from learning a fact or belief directly, but does little to prevent indirect communication. If an agent of a different sub-population (or even of the same population) had internet access, for example, it could learn the information contained on the web site. This agent could then pass on belief information to agents lacking web access via agent-to-agent communication. If the agent lacking information access changed its belief, it would be noted as a change in belief in the tables presented above. This would be noted as effectively a smaller change in the population, though in actuality the mechanism by which the effect occurred would be substantially different than in the case in which literacy and information access were not modeled.

Alternatively, the lack of literacy and internet access in one subpopulation may actually lead to modifications in other sub-populations. An agent lacking access, or who is affected by the truncation and misinformation effects of illiteracy, might be more likely to communicate with the promoter. This agent could then relay that information on to other agents (in the same socio-demographic sub-population or in different sub-populations), thereby increasing their knowledge, changing their belief, or modifying their behavior. If a sufficient amount of this between-population communication were to occur, it could lead agents in another socio-demographic group to change their knowledge belief, or behavior. This might also serve to decrease the difference between the agents in the affected socio-demographic group relative to other groups, leading to a smaller observed difference.

Following this logic, then it is possible to understand why the web intervention could be a powerful even in situations where it might not be expected to be. Consider, again, the breakdown by age in Section 6. The web intervention was effective for older agents, even though these agents were least likely to have direct access to the intervention. If a sufficient number of younger agents existed in the society, and the message of the web site was strong enough, it is likely that younger and middle-aged agents could pass on information to older agents. Simultaneously, younger older agents could serve as a conduit for the promoter's information, leading to additional knowledge, belief, and behavior in the younger socio-demographic group. This could lead the web intervention to have less of an effect on the younger population, as the younger population might otherwise have less information spread directly via the promoter. Thus, while the literacy and information access constraints would continue to affect older agents, the constraints do not prevent communication; depending upon the frequency, strength, and type of communication, the web site may appear to be more effective for one sub-population while simultaneously less effective for another sub-population.

Lastly, it is important to note that the results for the city were gathered using a city with a very stylized age, income, and education breakdown. While there were correlations implicitly in the stylized city data, by virtue of the fact that there were both more low education and low income agents, these correlations were the result of independent parameter assignment and the true underlying population was not always guaranteed to match patterns observed in an actual city. However, the data generated for the real cities was based on census data and was able to use correlations from that source. Past work with Construct has suggested that differences in simulation results may occur when the underlying model assumptions are changed and the correlations among agent parameters become stronger [1][2]. As discussed in Section 4 as well as several other points in this technical report, different results can occur if greater realism is added to the model. Therefore, again considering the age parameter, some of the results for the real and stylized city are similar – for instance, the oldest agents in cities such as Hartford and the stylized city, but others differ greatly, such as occur in the younger age brackets in most of the real cities modeled. This suggests that future work might be necessary to understand how the interrelated in socio-demographic parameters play a role in determining which types of agents are likely to talk with which kinds of interventions, and then to identify the types of intervention combinations which may be more or less effective.

10 Conclusion

This technical report described a series of virtual experiments conducted to try and isolate the effects of particular educational interventions and information access mechanisms on agents with specific socio-demographic attributes. It suggested that different interventions may be more effective for agents of certain socio-demographic characteristics, including age, income, and education. By modeling three different information access mechanisms – literacy, internet access, and newspaper readership – it was possible to better model some of the nuances that restrict the flow of information in a society. By breaking this down by agent traits, it was possible to examine the effect of the interventions on different groups in the society. While members of these different groups interact, and the effect of an access constraint on one group may be slightly modified due to interaction with members of another group that does not suffer from that access constraint as strongly, it is still possible to get an approximate idea of what kinds of behavior are possible in a society when different access constraints are enabled.

This technical report also sought to chronicle the parameters used in conducting these virtual experiments in order to provide documentation about the input decks used in this experiment. This action was not performed as effectively for past work, and this document sought to record a greater number of the parameters used. As many of the experimental parameters used in this deck have been used for some time, this document also serves as retroactive documentation for some of the input parameters used in past work. It is also likely that, barring new validation data, experimental overhaul, or feature addition, many of these parameters may be in use in future simulations.

It has been suggested that the interventions analyzed here be used in the creation of a stylized city generator in which it would be possible to examine the socio-demographic constraints of an arbitrary stylized city, then use the data gathered to try and predict the relative effectiveness of the intervention outcomes. The expectation would be that, if successful, such a methodology could be used to help predict the types of intervention combinations which might be useful on real cities of an arbitrary socio-demographic composition. This idea, and others, are excellent stepping stones for future virtual experiments and work with Construct.

11 Acknowledgements

The authors would like to thank Neal Altman and Mike Kowalchuck for their support and assistance with this research.

12 References

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

[2] Brian Hirshman, Michael Martin, Alaiksandr Birukou, Mike Bigrigg, and Kathleen Carley. The Impact of Educational Interventions on Real and Stylized Cities. Technical report, Carnegie Mellon University School of Computer Science, May 2008.

[3] Kathleen Carley. Communication technologies and their effect on cultural homogeneity, consensus, and the diffusion of new ideas. *Sociological Perspectives*, 38(4):547–571, 1995.

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

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

[6] Brian Hirshman and Kathleen Carley. Modeling information access in construct. Technical report, Carnegie Mellon University School of Computer Science, May 2008.

[7] Internal Revenue Service. *EITC for individuals*. March 2008.

[8] 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.

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

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

[11] Paul Erdos and Alfred Renyi. On random graphs i. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959.

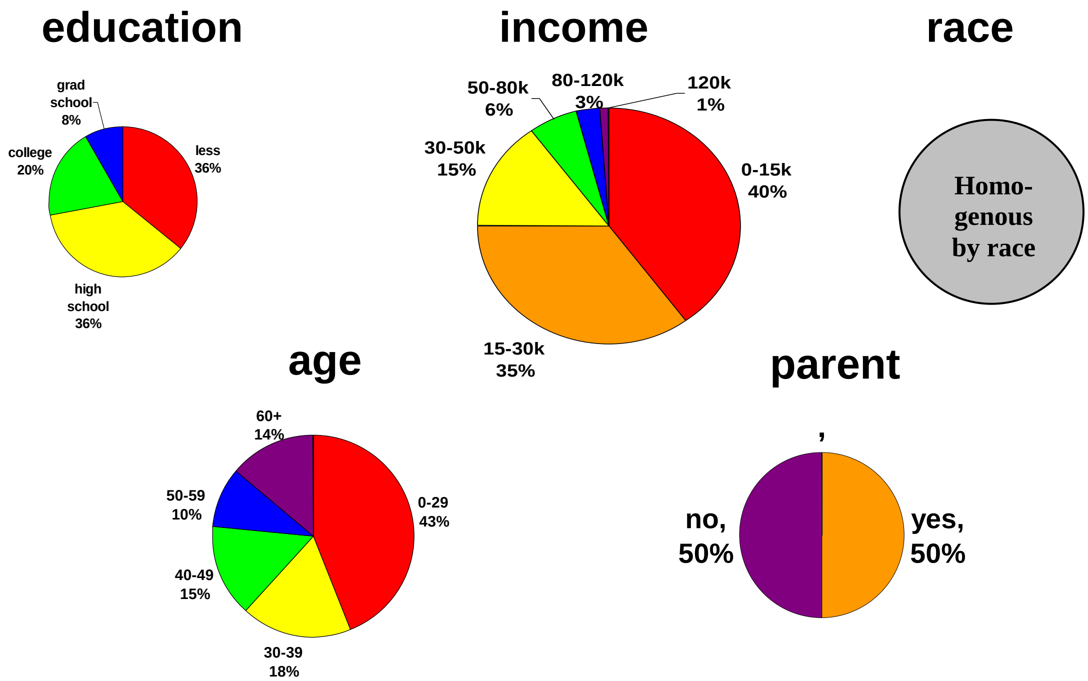
[12] Daniel Wegner. Transactive memory: A contemporary analysis of the group mind. In B. Mullen and G.R.Goethals, editors, *Theories of group behavior*, New York, 1987. Springer-Verlag.

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

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

Appendix A. Socio-demographic Breakdowns

Figure 1: Breakdown of stylized city by socio-demographic group



Appendix B. Intervention Effects by Age Group

Figure 2: Age Group by Intervention: Change in Know How.

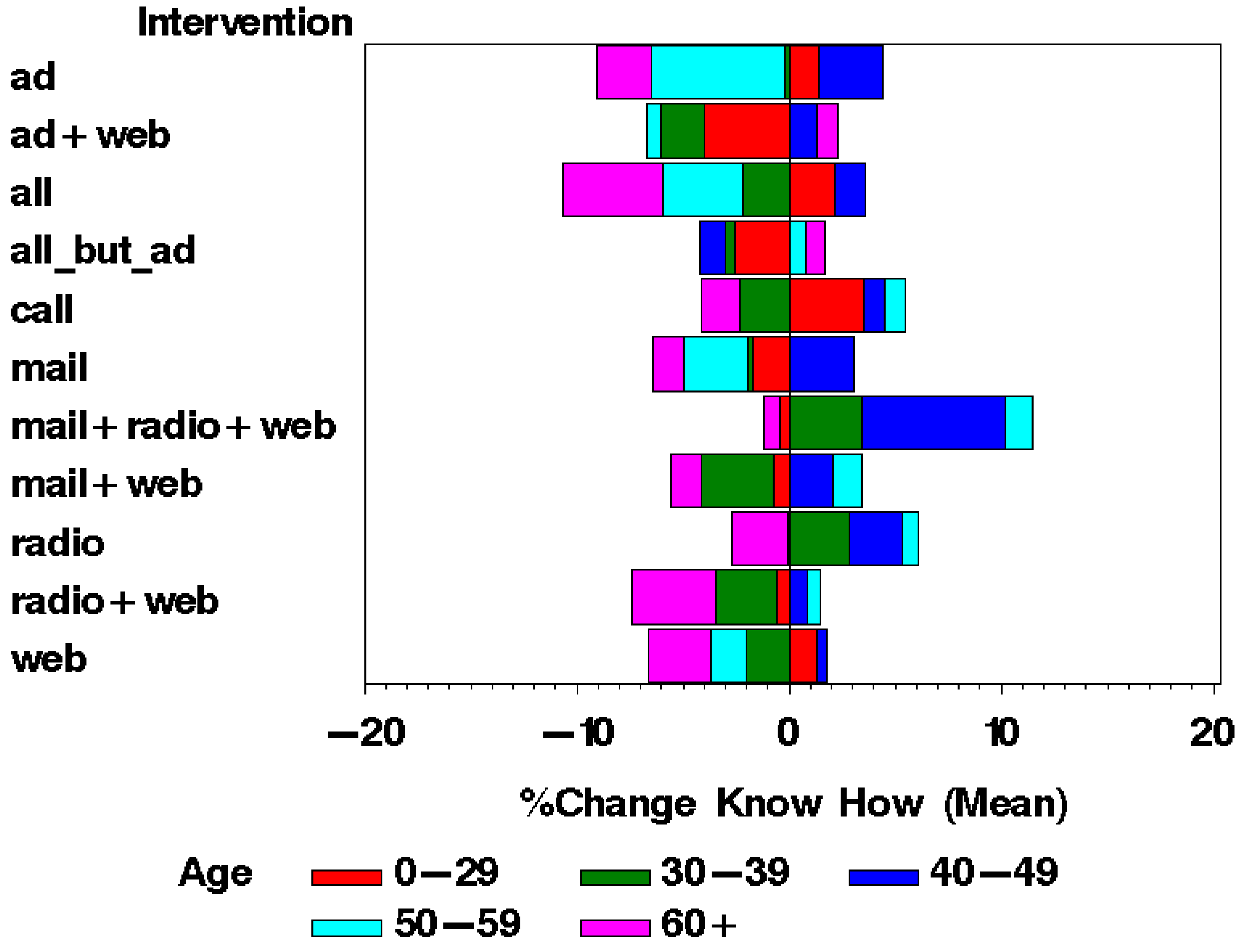


Figure 3: Age Group by Intervention: Change in Belief about Scheme.

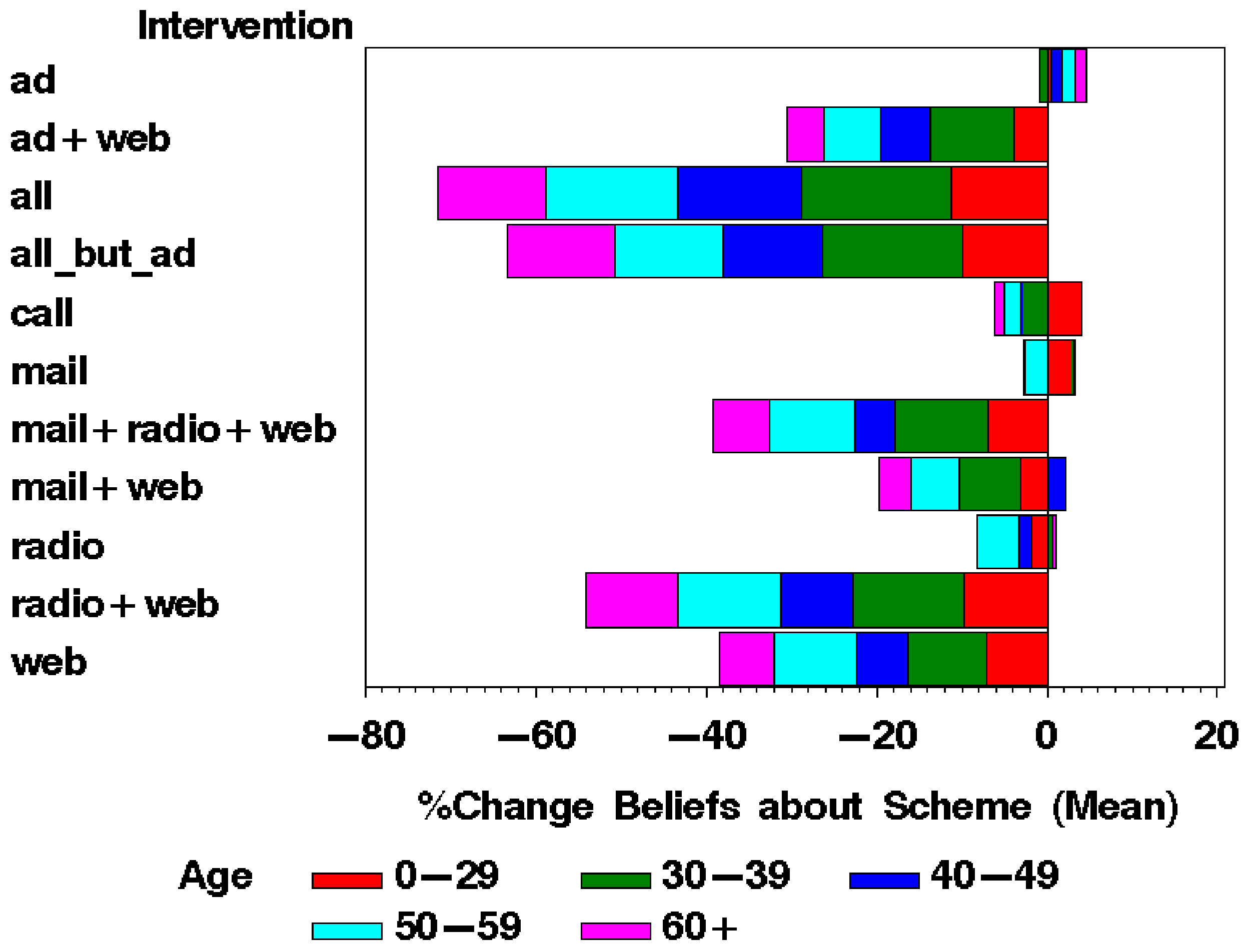
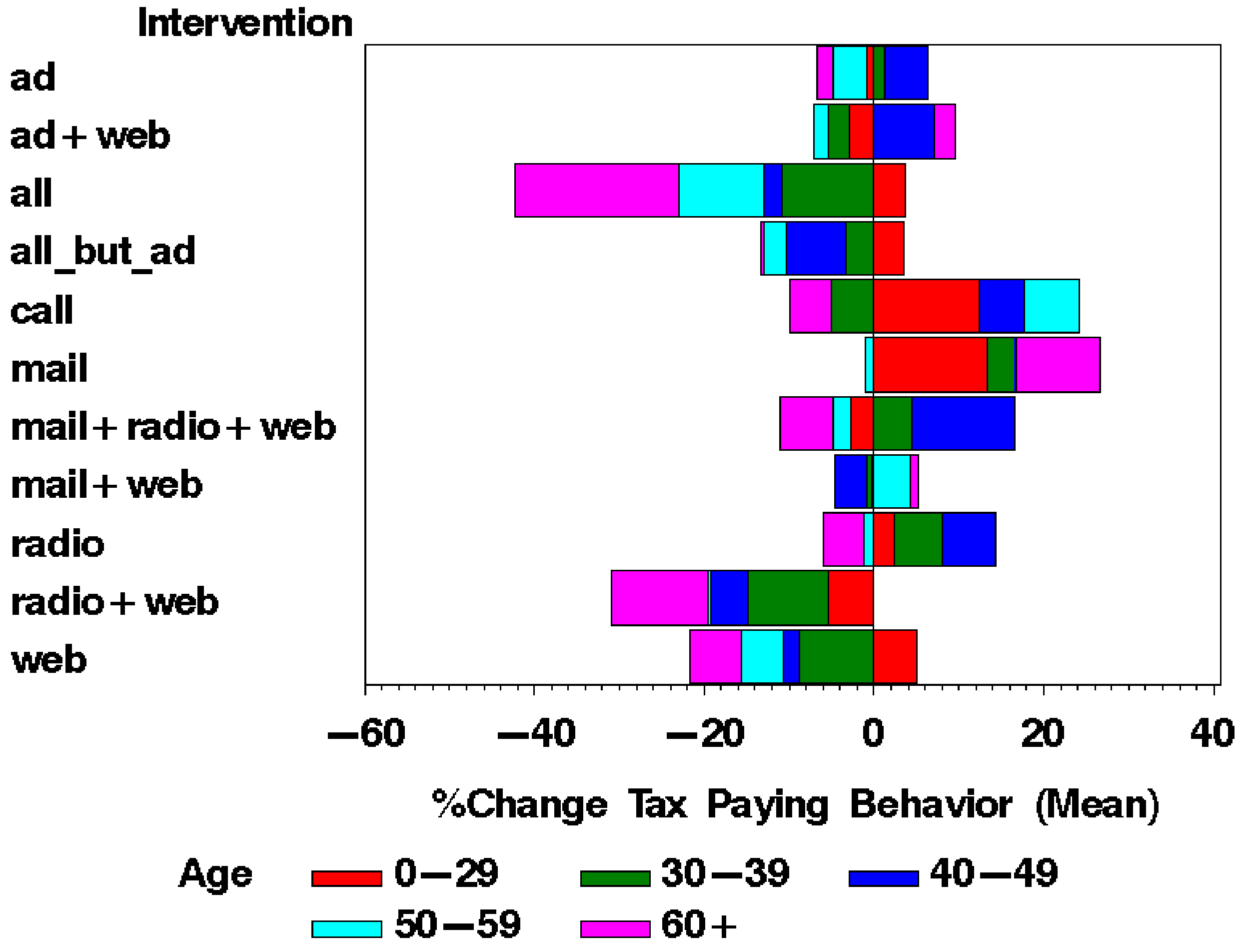


Figure 4: Age Group by Intervention: Change in Tax Paying Behavior.



Appendix C. Intervention Effects by Income Group

Figure 5: Income Group by Intervention: Change in Know How.

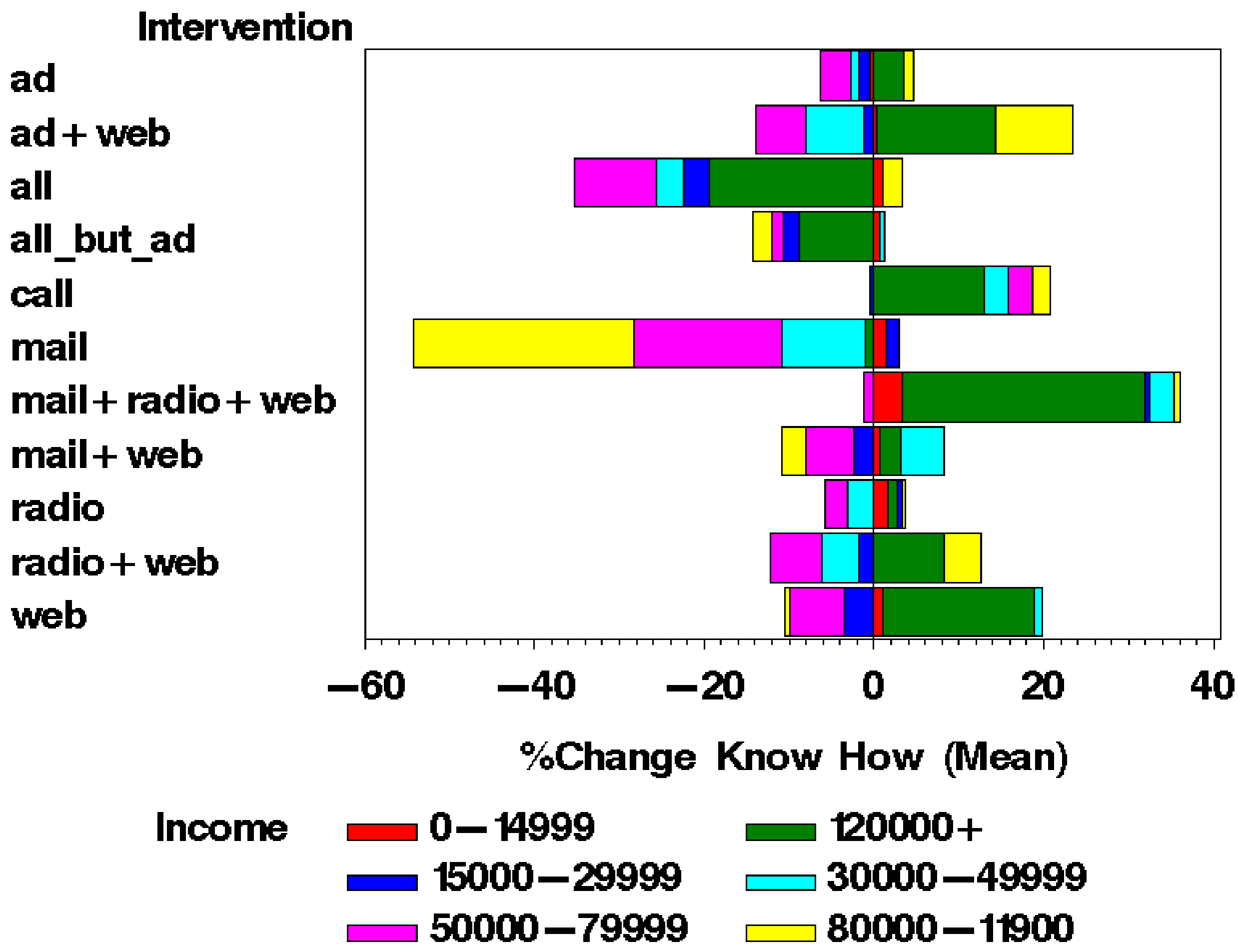


Figure 6: Income Group by Intervention: Change in Belief about Scheme.

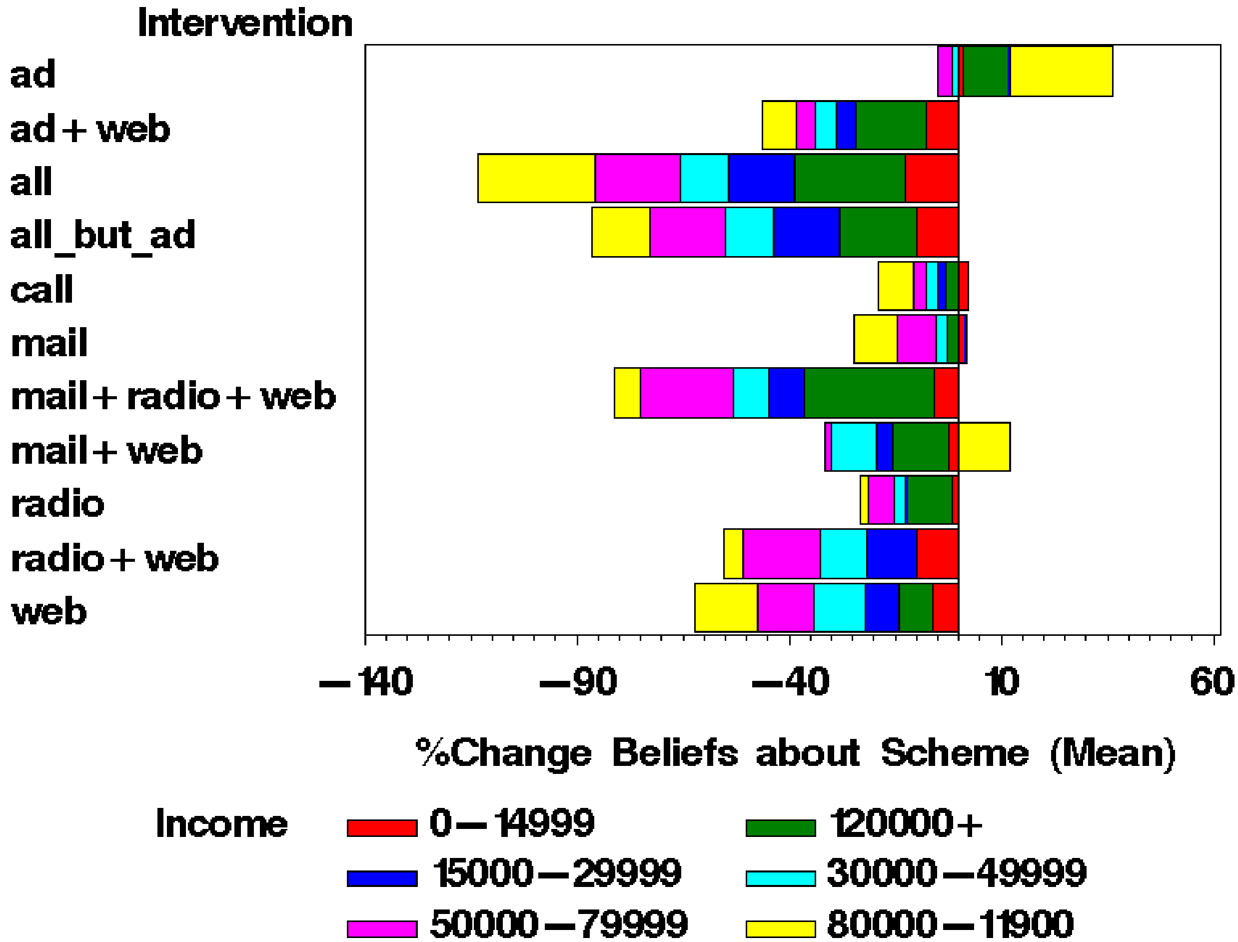
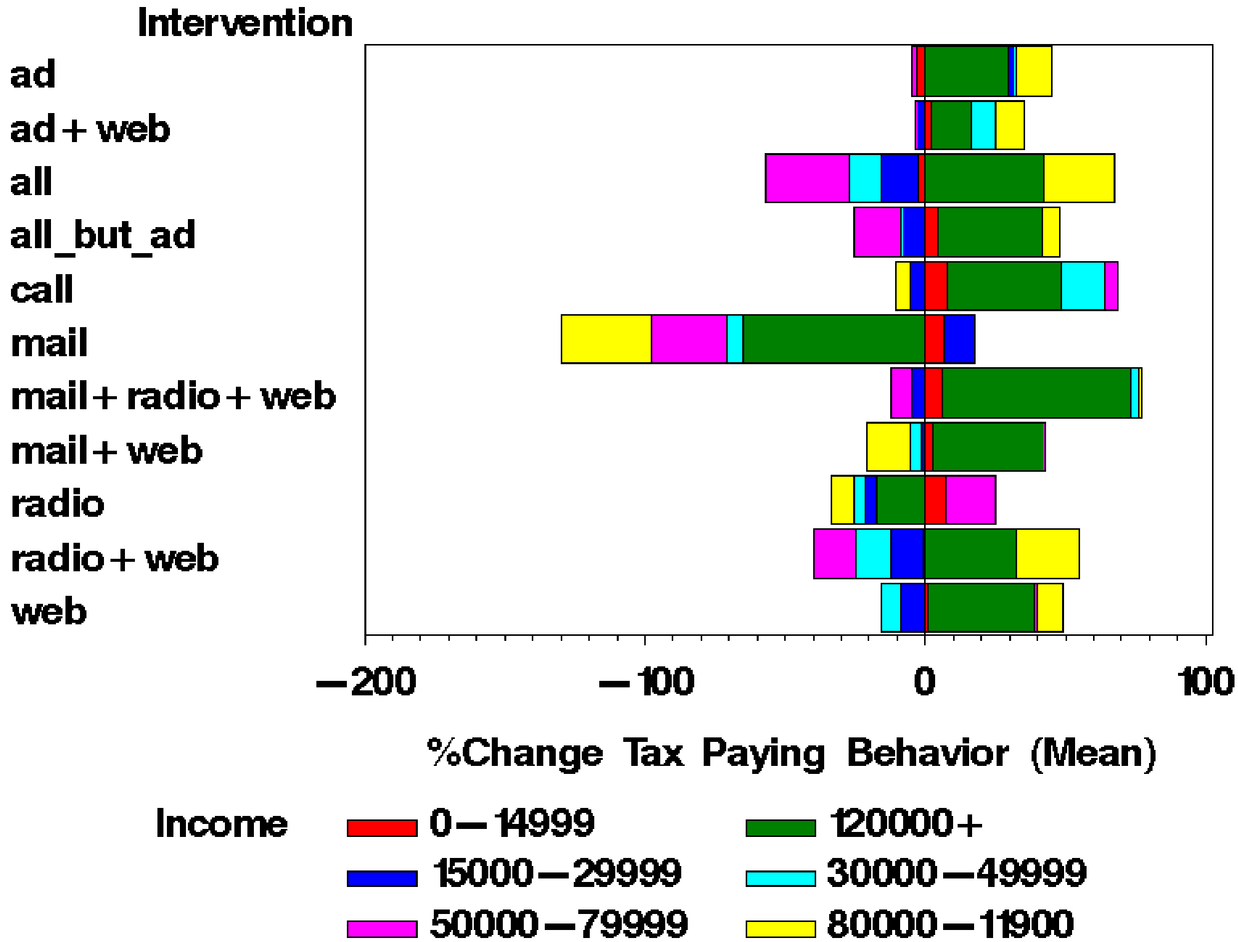


Figure 7: Income Group by Intervention: Change in Tax Paying Behavior.



Appendix D. Intervention Effects by Education

Figure 8: Education by Intervention: Change in Know How.

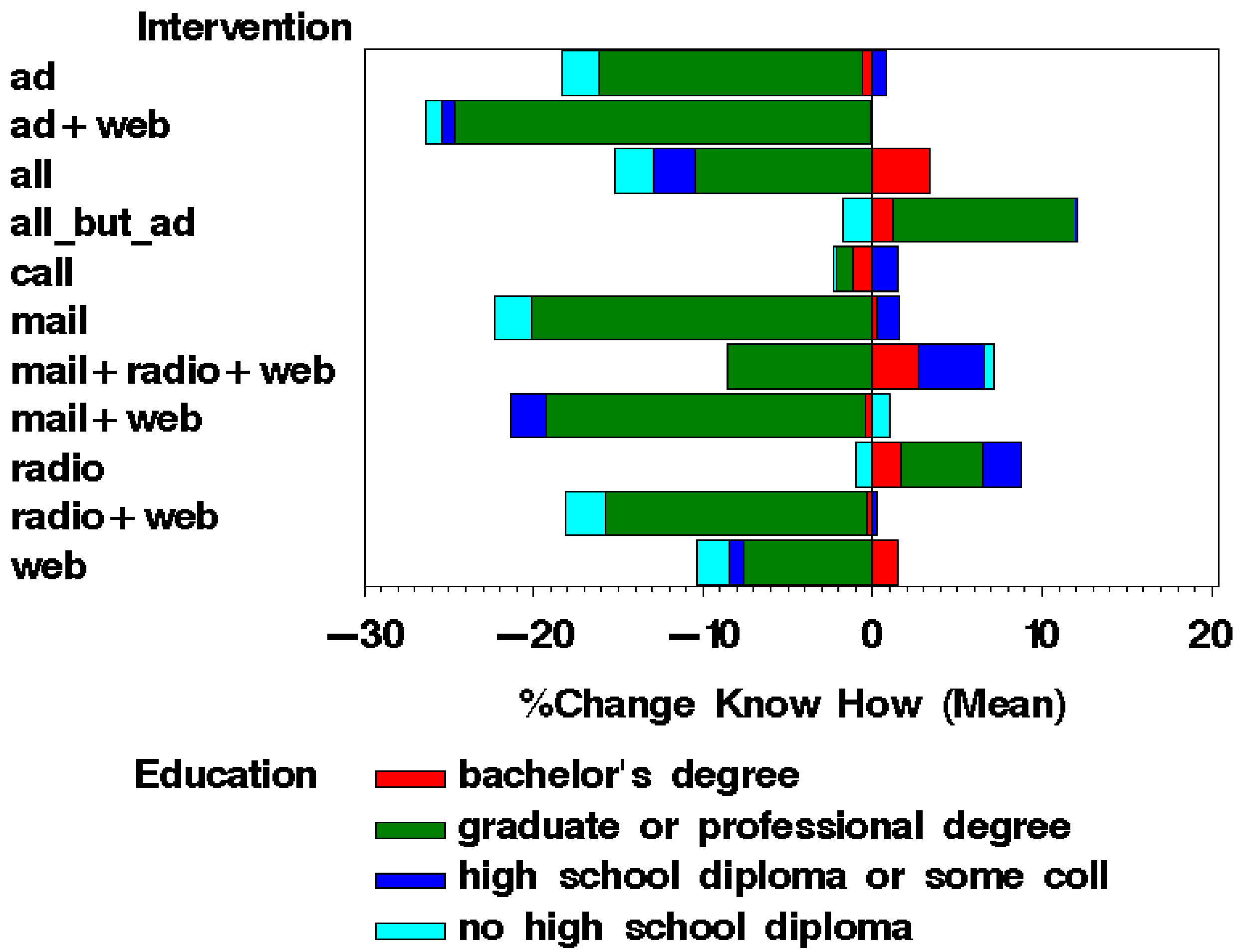


Figure 9: Education by Intervention: Change in Belief about Scheme.

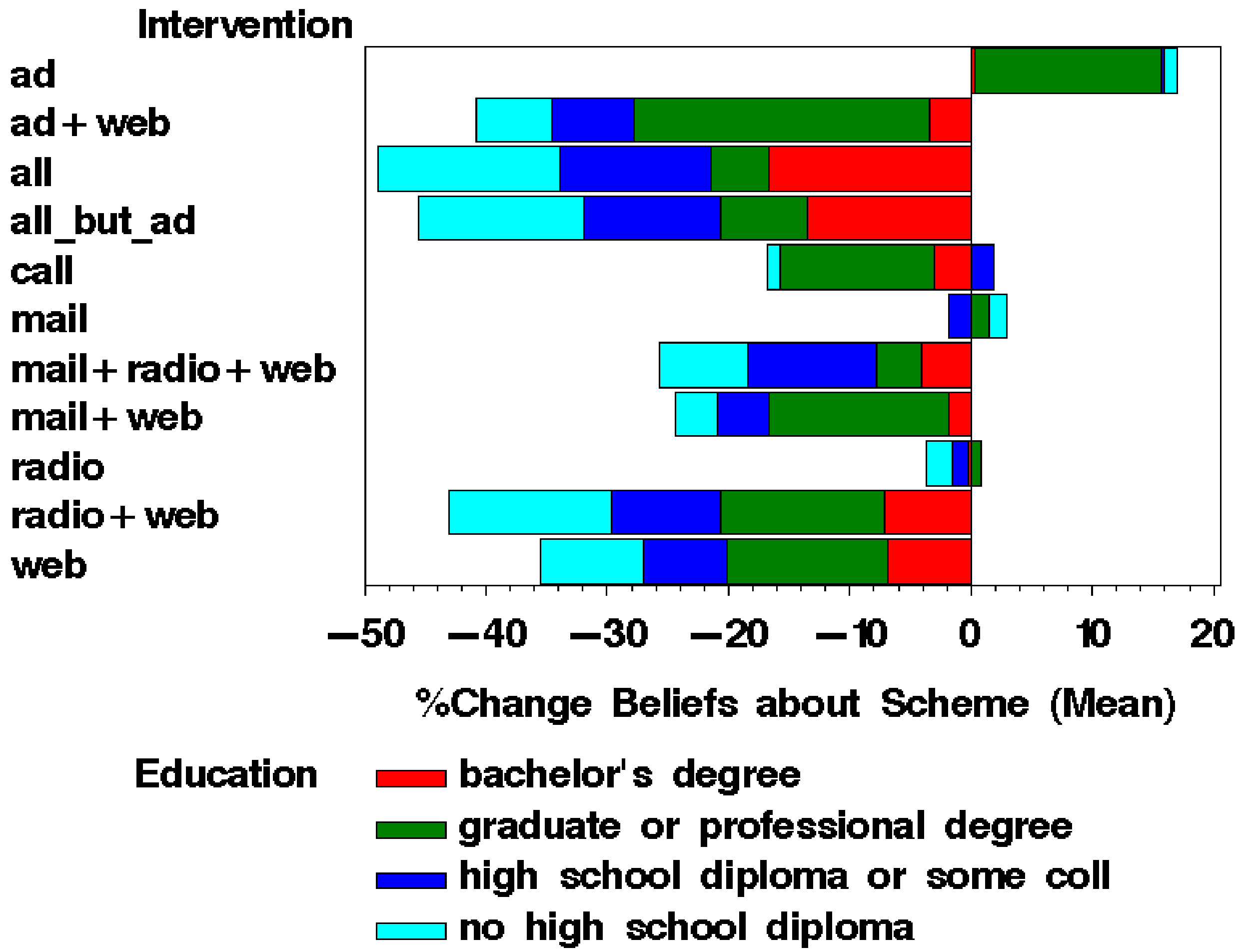


Figure 10: Education by Intervention: Change in Tax Paying Behavior.