



## Article

# An agent-based simulation of persistent inequalities in health behavior: Understanding the interdependent roles of segregation, clustering, and social influence



Brent A. Langellier\*

Department of Health Management and Policy, Dornsife School of Public Health, Drexel University, United States

## ARTICLE INFO

## Keywords:

Diet  
Health inequalities  
Complex systems  
Social influence  
Segregation

## ABSTRACT

Health inequalities are conspicuously persistent through time and often durable even in spite of interventions. In this study, I use agent-based simulation models (ABMs) to understand how the complex interrelationships between residential segregation, social network formation, group-level preferences, and social influence may contribute to this persistence. I use a more-stylized ABM, Bubblegum Village (BV), to understand how initial inequalities in bubblegum-chewing behaviors either endure, increase, or decrease over time given group-level differences in preferences, neighborhood-level barriers or facilitators of bubblegum chewing (e.g., access to bubblegum shops), and agents' preferences for segregation, homophily, and clustering (i.e., the 'tightness' of social networks). I further use BV to understand whether segregation and social network characteristics impact whether the effects of a bubblegum-reduction intervention that is very effective in the short term are durable over time, as well as to identify intervention strategies to reduce attenuation of the intervention effects. In addition to BV, I also present results from an ABM based on the distribution and social characteristics of the population in Philadelphia, PA. This model explores similar questions to BV, but examines racial/ethnic inequalities in soda consumption based on agents' social characteristics and baseline soda consumption probabilities informed by the 2007–2010 National Health and Nutrition Examination Survey. Collectively, the models suggest that residential segregation is a fundamental process for the production and persistence of health inequalities. The other major conclusion of the study is that, for behaviors that are subject to social influence and that cluster within social groups, interventions that are randomly-targeted to individuals with 'bad' behaviors will likely experience a large degree of recidivism to pre-intervention behaviors. In contrast, interventions that target multiple members of the same network, as well as multilevel interventions that include a neighborhood-level component, can reduce recidivism.

## 1. Introduction

## 1.1. Diet inequalities

Racial/ethnic minorities suffer from high rates of obesity and related chronic disease (Flegal, Carroll, Ogden, & Curtin, 2010). While physical (in)activity plays a part, these inequalities are at least partially driven by well-documented differences in diet (Go et al., 2013). High rates of obesity among Mexican Americans (~78%) may, in part, be driven by the high level of consumption of sugar-sweetened beverages (SSB) among that population (Flegal et al., 2010). Between-group differences in consumption of fast food and other foods away from home may also contribute to obesity disparities (Batis,

Hernandez-Barrera, Barquera, Rivera, & Popkin, 2011; Boone-Heinonen et al., 2011; Bowman, Gortmaker, Ebbeling, Pereira, & Ludwig, 2004). Improving diet quality is of critical public health importance, but doing so is difficult because diet is influenced by multilevel factors at the environmental (e.g., access to food resources), household (e.g., family composition, income), individual (e.g., educational attainment, preferences), and interpersonal (e.g., social influence) levels (Diez-Roux et al., 1999; Glanz, Sallis, Saelens, & Frank, 2005; Moore, Diez Roux, Nettleton, Jacobs, & Franco, 2009). Addressing inequalities is even more difficult, because racial/ethnic groups vary markedly in many of the characteristics (e.g., age, educational attainment, neighborhood) that most strongly influence diet. For example, the Latino population as a whole is younger, poorer,

\* Correspondence address: Department of Health Management and Policy, Dornsife School of Public Health, Drexel University, Room 356, Nesbitt Hall, 3215 Market Street, Philadelphia, PA 19104, United States.

E-mail address: [bal95@drexel.edu](mailto:bal95@drexel.edu).

<http://dx.doi.org/10.1016/j.ssmph.2016.10.006>

Received 6 July 2016; Received in revised form 21 September 2016; Accepted 13 October 2016

2352-8273/© 2016 The Author. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

and less well-educated than the White population. Thus, dietary inequalities could be driven by cultural differences (e.g., food preferences), environmental disparities in food access, differences in the group-level distribution of socio-demographic characteristics (i.e., population composition), or some combination thereof. Furthermore, due to the interdependent effects of residential, economic, and school-based segregation, individuals are likely to live, work, and play with others who share their racial/ethnic and economic characteristics. The influence exerted through homogeneous social networks may help consolidate health behaviors and exacerbate inequalities between groups.

### 1.2. Residential segregation and health

The high level of racial/ethnic and income-based segregation in American cities is well-documented (Massey & Denton, 1988, 1989, 1993). Segregation is associated with poorer educational outcomes (e.g., educational attainment and test scores) and highly concentrated poverty in minority neighborhoods (Card & Rothstein, 2007; Massey, Condran, & Denton, 1987; Massey & Denton, 1993; Massey & Fischer, 2000). This place-based intersection of race/ethnicity and socioeconomic status likely has broad implications for health inequalities, as income and education are widely accepted as fundamental social causes of disease (Link & Phelan, 1995). More proximate mechanisms through which segregation may produce inequalities is via disparities in access to neighborhood resources. This may include both social resources (e.g., neighborhood-level educational attainment or income) as well as physical resources (e.g., the food and physical activity environment, health care services). Most relevant to the current study, low-income and minority neighborhoods tend to have higher concentrations of fast food restaurants, small corner stores, and liquor stores but decreased access to comprehensive supermarkets (Larson, Story, & Nelson, 2009; Morland, Wing, Diez Roux, & Poole, 2002). Several studies have observed a relationship between the local food environment and dietary or weight outcomes, although most of these studies are based on cross-sectional data and there is no consensus on whether this relationship is causal (de Vet, de Ridder, & de Wit, 2011; Holsten, 2009). Among the limited longitudinal studies in this area, Boone-Heinonen and colleagues (2011) found that neighborhood density of fast food restaurants is related to fast food consumption among low-income adults (Boone-Heinonen et al., 2011).

### 1.3. Clustering of health behaviors and outcomes

Health behaviors and outcomes tend to cluster within social networks. This has most prominently been documented in a series of studies by Christakis, Fowler, and others regarding such diverse health issues as obesity, smoking, alcohol consumption, and happiness (Christakis & Fowler, 2007, 2008; Fowler & Christakis, 2008; Rosenquist, Murabito, Fowler, & Christakis, 2010). As described by Shoham and colleagues (2012), there are three explanations that could each independently produce this type of clustering: (1) individuals could be attracted to others that share their health behaviors and outcomes (i.e., homophily), (2) individuals within a social network may share exposure to environmental, social, cultural or other factors that shape behavior (i.e., common causes), and (3) an individual's behavior may be influenced by the behavior of members of their social networks (i.e., social influence).

Assessing the extent to which each of these three mechanisms contribute to clustering of health behaviors and outcomes is important for understanding disease dynamics at the population level. In the absence of other clustering mechanisms, homophily in and of itself would have little or no impact on inequalities because social network formation would not change within- or between-group distributions of health outcomes. The 'common causes' clustering hypothesis is largely consistent with the social determinants and health disparities literature

(e.g., the fundamental causes of disease hypothesis of Link and Phelan (1995)). In general, this literature asserts that social gradients in the distribution of 'upstream' factors (e.g., income, education, power) are the key drivers of disparities in 'downstream' health factors. The social influence hypothesis is consistent with behavioral theory, most prominently Bandura's Social Cognitive Theory (Bandura, 1986). Furthermore, as described in a review of social influence and obesity conducted by Cunningham, Vaquera, Maturo, and Narayan (2012), several longitudinal studies have provided evidence in support of social influence.

If social influence is a driver of clustering, there would be important implications for health inequalities and health promotion interventions. Residential, school-based, and other forms of segregation may result in homogeneous networks across factors like race/ethnicity and income. Against this backdrop of homogeneous social networks, social influence that consolidates health behaviors within groups would likely exacerbate between-group inequalities. A further consequence of social influence may be positive 'spillover' effects on the friends and family of those who participate in an intervention (Trogon & Allaire, 2014). Conversely, social influence may dampen intervention effects over time. Weight loss interventions suggest that maintenance of intervention effects is a challenge for many participants (Wing & Phelan, 2005; Wing, Tate, Gorin, Raynor, & Fava 2006). Such interventions might achieve positive behavior change, but social influence from friends and family may 'pull' the participant back to the pre-intervention levels exhibited by his/her social network. If social influence is important, enrolling multiple members of the same social network may be an effective strategy to promote maintenance of intervention effects.

### 1.4. Complex systems research in diet

A growing body of research has used complex systems methods like agent-based modeling and system dynamics modeling to understand social processes that impact diet and other health behaviors. In brief, complex systems methods are simulation-based approaches that allow researchers to examine the potential impacts of feedback loops, non-linear effects, adaptation, and other dynamic processes that change over the course of time (Mabry, Marcus, Clark, Leischow, & Méndez, 2010). Health researchers, for example, have used agent-based models (ABMs) to examine adaptive behaviors including the spread of health behaviors through social networks via social influence (Hammond & Ornstein, 2014; Orr, Galea, Riddle, & Kaplan, 2014), adaptation of food stores in response to purchasing patterns in a given neighborhood (Auchincloss, Riolo, Brown, Cook, & Diez Roux, 2011), and the impact of past experiences on future behavior (Yang, Roux, Auchincloss, Rodriguez, & Brown, 2011). These mechanistic models seek to identify and understand features of complex systems (e.g., feedback loops, social influence) that contribute to population-level patterns in health behaviors and outcomes. A further body of literature uses complex systems models to identify leverage points for interventions to help practitioners and policymakers decide between one or a combination of interventions (Widener, Metcalf, & Bar-Yam, 2013; Zhang, Giabbanelli, Arah, & Zimmerman, 2014).

### 1.5. The present study

In this study, I use agent-based models (ABMs) to explore group inequalities in health behaviors. I have two primary objectives: First, I examine the potential implications of residential segregation, social network formation, and social influence for persistence of group inequalities. Second, I explore the durability of interventions to improve health behaviors in the presence of social influence. Broadly, I focus on three types of interventions: those that target random individuals among the minority population, those that target multiple members of the same social network, and those that target the environment.

To address these objectives, I use a highly stylized model to explore the implications that extreme levels of segregation, homophily, and clustering have on durability of health inequalities. I frame the stylized model within the context of understanding inequalities in bubblegum chewing behavior, a metaphor for health behaviors that vary between groups and are subject to social and environmental influences. As a supplement to the stylized model, I present an ABM of soda consumption in a model environment that approximates the socio-demographic characteristics and spatial distribution of the population of Philadelphia, PA. Baseline (i.e., prior to social influence) soda consumption preferences of each agent are ‘anchored’ to parameters derived from a national diet study. The intent of this model is to explore whether implications from the stylized model are salient to an important health behavior given the high levels of residential segregation and plausible between-group differences in health behaviors that might be observed in an American city.

## 2. Material and methods

### 2.1. Overview of model environments

Below I overview the simulation models used in this study. Further details regarding the implementation of the models, key parameters, assumptions, and sensitivity analyses can be found in [Appendix A](#).

I use two agent-based simulation models to explore the implications of residential segregation, social network characteristics, and social influence on health inequalities. The first model is Bubblegum Village (BV), a stylized model of red and green agents. I use BV to explore the implications that extreme (i.e., very high or very low) levels of segregation, homophily, and clustering may have on the durability of health inequalities. Each agent in BV lives in a neighborhood, has a social network of between three and five friends drawn from their neighborhood and adjacent neighborhoods, and makes a daily decision regarding whether or not to chew bubblegum. The model allows for varying levels of segregation preferences, homophily preferences (e.g., the preference of a green person to be friends with other green people), and clustering (i.e., increase in the proportion of ‘friends of friends’). Upon initialization, population-level bubblegum preferences differ between the red and green groups and neighborhood-level barriers and facilitators differ between majority-green and majority-red neighborhoods.

BV is a metaphor for complex systems characterized by the following: (1) existence of majority and minority groups, (2) group-level inequalities in initial preferences, (3) inequalities in neighborhood-level influence, (4) social influence. Examples of systems that generally meet these criteria might include diet, physical activity, or substance use behaviors in most American cities. The purpose of BV is to explore how residential segregation, neighborhood-level influences, social influence, and initial between-group differences in behavioral preferences interact to produce inequalities between majority and minority groups.

I also use an ABM of sugar sweetened beverage (SSB) consumption in Philadelphia, PA to examine whether social network characteristics have implications for production and persistence of health inequalities given observed levels of segregation in an American city and plausible between-group differences in an important health behavior. The model is based in GIS space, with a synthetic population of agent-individuals in size proportionate to 2% of the Philadelphia population. Each individual is assigned age, gender, marital status, educational attainment, income, and race/ethnicity (white, black, Latino, and Asian) in proportion to census tract-level data from the 2010–2014 American Community Survey and the 2010 Decennial Census. Children are assigned to the nearest elementary/middle or high school, depending upon their age, and adults are assigned to a random workplace. The location of schools and the healthfulness of the neighborhood food environment are based on observed data from Philadelphia.

### 2.2. Important model processes

Each model includes a number of important processes related to residential segregation, social network formation, social influence, and health behaviors. Full details regarding these processes can be found in [Appendix A](#). In brief, agents in BV first self-select into neighborhoods based on preferences regarding the extent to which they prefer to live in a neighborhood where a majority of other residents are of the same color (i.e., red agents prefer majority-red neighborhoods). This segregation preference varies across iterations of the model, with the intent of producing environments with no segregation up to very segregated environments. Agents in BV also have preferences that impact the formation of social networks. These include preferences for homophily and clustering. These preferences also vary across iterations, but produce social networks with high or low levels of segregation and clustering.

The main outcome in BV is the frequency with which red and green agents chew bubblegum. This frequency is a function of an individual-level attitude and a neighborhood-level influence that reflects barriers or facilitators to chewing. At model initialization, I designate 75% of red individuals are designated as frequent chewers and 25% as infrequent chewers; among greens the reverse is true. Similarly, I assign 75% of red-majority neighborhoods as being facilitators of bubblegum chewing, while only 25% of green-majority neighborhoods are facilitators. These baseline inequities are consistent with my intent to use BV to explore systems characterized by existence of majority and minority groups, group-level inequalities in initial preferences, and inequalities in neighborhood-level influence.

The final important process in BV is the influence of social network members on bubblegum chewing behavior. Social influence operates via feedback loops adapted from Yang (2011). In brief, these feedbacks update each an agent's attitudes based on the behaviors of ‘friends’. Agents periodically observe the behaviors of their friends and update their own attitudes to conform. For example, if a red agent only chews bubblegum once per week but his/her friends chew bubblegum an average of five days per week, that agent's attitudes towards bubblegum will increase a small amount (making him/her more likely to chew bubblegum during the subsequent period of time).

Processes in the Philadelphia model are similar to those in BV, with a few notable exceptions. First, the distribution of the population across Philadelphia neighborhoods, including racial/ethnic distribution, is based on observed data from the U.S. Census. As such, the level of segregation remains the same across iterations of the model. In the Philadelphia model, individuals each have a ‘family/ household’ network in addition to a network of ‘friends’. Family networks consist of everyone that lives in the same housing unit. Friendship selection is similar to the BV model, except that friendship selection is based on varying preference for racial/ethnic homophily and clustering, as well as preference for similar age and gender. Similarly, social influence in the Philadelphia model is based on influence of an agent's network of friends, as well as a separate influence of the family network. Finally, each agent in the Philadelphia model is assigned a baseline probability of consuming at least one SSB on a given day. This probability is then updated as each run of the model progresses due to social influence. I assign baseline probabilities based on agent's socio-demographic and neighborhood characteristics, as well as weights derived primarily from the National Health and Nutrition Examination Survey. In brief, the weights convey the importance of each characteristic (e.g., household income) for SSB consumption. Please see [Table A.1](#) for a list of all parameter values.

### 2.3. Intervention scenarios

I include three intervention scenarios in both the BV and Philadelphia models. The interventions ‘improve’ health behaviors among a targeted group of agents from the minority groups in each

model. Intervention 1 improves attitudes ( $A_{ij}$ ) among a set of randomly-selected agent (either from the red group in BV or the Latino population in the Philadelphia model). The intervention includes 99 total participants in both BV and the Philadelphia model, and seeks specifically to identify agents with ‘bad’ behaviors (e.g., bubblegum chewing above a certain threshold) to recruit into the intervention. If there are not enough agents with ‘bad’ behaviors to fill the intervention, the model then selects agents at random for participation. Intervention 2 improves behaviors among the same number of agents, but is ‘network based.’ Specifically, the intervention includes a number of ‘primary participants’ as well as two members of each primary participant’s social network (total  $n=99$ ). In Interventions 1 and 2, the attitudes of intervention participants are ‘improved’ to a value determined via a random draw from a uniform distribution from (0.05, 0.15). Intervention 3 is an ‘environmental’ intervention that addresses neighborhood-level facilitators of either bubblegum chewing or SSB consumption. In BV, the intervention reduces the influence of ‘facilitator’ neighborhoods by an amount drawn from a uniform distribution:  $U(0.15, 0.05)$ . Since bubblegum chewing is a function of neighborhood influence, this effectively reduces bubblegum-chewing probabilities of all people in ‘bad’ neighborhoods, should have a greater impact on the red population than the green population, and should thus reduce inequalities. In the Philadelphia model, I improve the food environment in all census tracts with  $\geq 40\%$  Latino residents and a modified Retail Food Environment (mRFEI) score above a specific threshold. In addition to these three intervention scenarios, I also test ‘multilevel’ interventions that simultaneously include the individual-level interventions and the food environment intervention. All interventions are initiated after a ‘warm up’ period in which individual preferences reach stability (i.e., individual preferences are no longer changing due to social influence). I determined this warm up period via visual inspection of a running means plot of bubblegum chewing during multiple runs of the baseline scenario.

2.4. Model iterations and outcomes

To understand how different combinations of segregation and social network formation impact the persistence of inequalities, I use a factorial design that includes all possible combinations of strong vs. no segregation preferences (BV model only), strong vs. no homophily preferences, and strong vs. no preferences for clustering. To examine the durability of intervention effects, I examine combinations of individual-level interventions (no intervention, random intervention, network-based intervention) and the neighborhood interventions. I run these interventions in two environments: one with no preferences for segregation or network formation, and a second with high preferences for segregation, clustering, and homophily.

To assess differences across different iteration of the models, I collect mean weekly (i.e., every seven time steps) bubblegum consumption for the red and green populations in BV and mean number of days in the previous week with at least one SSB consumed, stratified by race/ethnicity. I average these outcomes across 25 runs (i.e., a single ‘run’ of the model) for BV and 20 runs of the Philadelphia model.

3. Results

3.1. Bubblegum Village

I present descriptive statistics regarding social network characteristics and persistence of bubblegum chewing inequalities in BV in Table 1. Scenario 1 is a ‘random’ model, with no preferences for segregation, clustering or homophily. In general, simulated levels of segregation, and homophily in this scenario largely reflect the fact that 70% of the population is green and 30% is red. Because the population is randomly distributed across neighborhoods, almost everyone in the model lives in a ‘majority-green’ neighborhood. Because friendship

**Table 1**  
Social network characteristics and persistence of bubblegum chewing by agent color in Bubblegum Village.

	Scenario Number							
	S1	S2	S3	S4	S5	S6	S7	S8
<b>Model Parameters</b>								
Segregation Pref.	0	0	0	10	0	10	10	10
Clustering Pref.	0	10	0	0	10	10	0	10
Homoph Pref.	0	0	10	0	10	0	10	10
<b>Segregation</b>								
Red in Red Neighb. (%)	6.2	4.4	5.6	70.8	5.1	69.3	70.5	68.4
Green in Green Neighb. (%)	97.5	98.2	97.8	88.8	98.0	89.9	89.4	89.3
<b>Homophily in Friendships (%)</b>								
Red	30.8	30.7	86.5	47.2	86.2	46.5	89.5	88.6
Green	70.2	70.6	94.3	77.2	94.2	77.6	95.6	95.2
<b>Clustering (%)</b>								
Red	18.0	56.1	27.2	18.3	35.4	56.0	25.5	33.4
<b>Weekly Bubblegum Chewing</b>								
Red – Begin	4.43	4.45	4.47	4.87	4.44	4.85	4.84	4.89
Red – End	1.70	1.67	2.70	3.52	2.63	3.50	4.62	4.57
Green – Begin	1.79	1.80	1.81	1.87	1.79	1.84	1.85	1.87
Green – End	1.42	1.40	1.16	1.84	1.12	1.79	1.32	1.45

Notes: Homophily % refers to the percent of all friendship ties in the model that are between color-concordant individuals. Clustering % refers to the percent of all friendship ties where an ego and connected alter each share a friendship tie with a third individual (i.e., % friend-of-a-friend). ‘Begin’ outcomes are calculated at model time step 7; ‘end’ outcomes are calculated at time step 497 (70 weeks later).

formation is random, an average of about 70% of the friends of green and red agents are green (i.e., ~70% homophily among green agents, ~30% homophily among red agents). Clustering (i.e., % of ties where the ego and alter share a third friend in common) is relatively low in the random model (18%).

In contrast, Scenario 8 includes strong preferences for segregation, clustering, and homophily. As might be expected, a much higher proportion of red people lived in majority-red neighborhoods in this scenario (68%) but slightly fewer green people lived in green-majority neighborhoods (89%). Nearly nine in ten friendships among red people were with other red people, and 95% of ties among green people were with other green people. Clustering increased to 33%.

Scenarios 2–7 represent ‘intermediate’ scenarios where there is some combination of preferences for segregation, clustering and homophily, but not strong preferences for all three. In general, segregation and network characteristics correspond with the agent preferences for these scenarios. Notably, clustering is considerably higher in Scenario 2 (clustering preferences only) and Scenario 6 (clustering and segregation preferences) than scenarios that include both clustering and homophily preferences. Similarly, homophily in friendships among both red and green agents is highest in scenarios where there are no preferences for clustering. This is likely because preferences for clustering and homophily likely have offsetting effects (e.g., the impact of a preference for clustering has a greater effect when it is the only criterion with which agents select friends).

Table 1 also includes mean weekly bubblegum chewing among red and green agents, at both initialization of the model and after 70 weeks have passed. As shown in Table 1, mean bubblegum chewing at initialization is about 1.8 days per week among green agents. Bubblegum chewing among green agents decreases over the course of the 70 simulated weeks in all scenarios. This is likely a function of model conditions. Specifically, since 70% of agents in the model are green, green agents in all scenarios will have mostly green friends. Furthermore, since 75% of green agents are infrequent chewers of bubblegum, most of the green agents who are frequent chewers will have friends that mostly chew bubblegum infrequently. Due to the

effects of social influence, therefore, many green agents that start out as frequent chewers will be ‘pulled’ towards the behaviors of their friends (who are likely to be infrequent chewers) as the model progresses. As might be expected, this effect (i.e., the consolidation of ‘good behavior’ among the green agents) is greatest in scenarios with homophily preferences. This is likely because, in these scenarios, green agents who are frequent chewers are less likely to have friends who are red agents, and thus are likely to have a greater proportion of friends who are infrequent chewers.

The effects of segregation, clustering, and homophily are even more drastic among the red agents. For red agents, mean bubblegum chewing at initialization is about 4.5 in scenarios without segregation and 4.9 in scenarios with segregation. This difference is likely because most majority-red neighborhoods facilitate bubblegum chewing. In scenarios without segregation, however, there are very few majority-red neighborhoods because most agents in the model are green. In scenarios with segregation there are more red neighborhoods and thus more red agents are influenced by ‘unhealthy’ neighborhoods.

In the random scenario, there is a marked decrease in red-green inequalities over time. By the end of the run, average bubblegum chewing among the green population is 1.42 days per week, compared to 1.70 among the red population. There is a similar pattern in Scenario 2, which includes clustering preferences but no homophily or segregation preferences. There are also large, albeit less dramatic, decreases in inequalities in Scenarios 3 and 5, which include homophily preferences but no segregation preferences. In contrast, inequalities largely persist in scenarios with preferences for segregation (4, 6, 7, 8). In the two scenarios with preferences for both segregation and homophily (7 and 8), there is very little change over time in the magnitude of inequalities.

I also used BV to explore whether social network characteristics may have implications for interventions to improve health behaviors among minority groups and thus reduce inequalities. Table 2 reports the mean bubblegum chewing of red agents that participated in four interventions (described in ‘Methods’): (1) random individual intervention, no neighborhood intervention, (2) random individual intervention with a neighborhood intervention, (3) network intervention, no neighborhood intervention, and (4) network intervention and neighborhood intervention. I present intervention effects for both Scenario 1 (no segregation, homophily, or clustering preferences) and Scenario 8 (preferences for all three).

Two things are clear from Table 2: first, the need for an intervention to reduce inequalities is much greater in Scenario 8 than Scenario 1. Pre-intervention bubblegum chewing is about 1.9 days/week in the

**Table 2**

Social network characteristics and durability of effects of an intervention to reduce bubblegum chewing among red people in Bubblegum Village.

	Scenario 1				Scenario 8			
<b>Model Parameters</b>								
Segregation Pref.	0				10			
Clustering Pref.	0				10			
Homophily Pref.	0				10			
	<b>Intervention Type</b>				<b>Intervention Type</b>			
	R/NF	R/F	N/NF	N/F	R/NF	R/F	N/NF	N/F
<b>Weekly Bubblegum Chewing</b>								
Pre-Intervention	1.80	1.92	1.89	1.92	6.37	6.38	6.13	6.15
Post-Intervention	0.56	0.38	0.59	0.39	1.53	0.90	1.49	0.92
End	1.09	0.50	1.13	0.47	5.48	2.98	5.11	2.79

Notes: Pre-intervention is at time step 490; post-intervention is at 511; end is at 1001. R=randomly-targeted intervention; N=network-based intervention; NF = no food environment intervention; F=food environment intervention.

random scenario, compared to over 6 days/week in the scenario with preferences for segregation, homophily, and clustering. This is largely because, in the random model, almost all red agents have mostly green friends. As a result of social influence, the bubblegum chewing behaviors of these red agents comes to closely resemble the ‘good’ patterns exhibited by the majority-green agents. As a result, there are few red agents with ‘bad’ behaviors to participate in the intervention so most intervention participants have relatively good behaviors to begin with.

The second implication of results in Table 2 is that there are fairly substantial differences in the durability of intervention effects based on type of intervention. Specifically, in the model with preferences for segregation, homophily, and clustering, participants in the ‘random’ intervention with no neighborhood intervention demonstrated nearly full recidivism to ‘pre-intervention’ bubblegum chewing levels over the 70-week period. This level of recidivism was slightly less in the network-based intervention, and much less in scenarios that included the neighborhood intervention. This suggests that, for behaviors subject to social influence, network-based interventions may have minor benefits relative to randomly-targeted interventions. However, neighborhood-level interventions or other interventions that target large segments of the population will likely have the most durable impact.

**3.2. Philadelphia model**

Table 3 includes social network characteristics for eight simulation scenarios in the Philadelphia model, which range from no preferences for homophily or clustering to preferences for both. Similar to BV, simulated levels of homophily and clustering vary greatly based on preferences. Homophily among the white and black populations, the two largest racial/ethnic groups in Philadelphia, is highest in models both without (56–59%) and with (91–93%) preferences for homophily. Homophily in all scenarios is lowest for Asians, the smallest and most diffuse group.

Table 3 also shows mean soda consumption by race/ethnicity over

**Table 3**

Social network characteristics and soda consumption by race/ethnicity in Philadelphia.

	S1	S2	S3	S4	S5	S6	S7	S8
<b>Model Parameters</b>								
Homophily Pref.	0	0	0	0	25	25	25	25
Clustering Pref.	0	0	25	25	0	0	25	25
<b>Family Influence</b>	No	Yes	No	Yes	No	Yes	No	Yes
<b>Racial Homophily in Social Ties (%)</b>								
White	58.7	58.4	57.2	57.1	92.3	92.4	91.4	91.3
Black	57.9	58.2	56.4	56.7	92.5	92.7	91.7	91.8
Latino	25.4	25.4	24.0	24.2	80.6	80.6	78.0	77.8
Asian	11.3	11.4	11.5	11.3	69.7	69.5	65.7	64.5
<b>Clustering (%)</b>	26.6	26.8	75.6	75.6	36.0	35.8	51.2	51.3
<b>Weekly Soda Consumption</b>								
White – Begin	2.84	2.84	2.84	2.84	2.84	2.84	2.83	2.84
White – End	3.08	3.24	3.06	3.21	3.09	3.25	3.07	3.25
Black – Begin	3.09	3.07	3.10	3.10	3.09	3.08	3.09	3.10
Black – End	3.10	3.07	3.07	3.04	3.09	3.03	3.09	3.04
Latino – Begin	3.57	3.55	3.57	3.55	3.55	3.57	3.55	3.56
Latino – End	3.40	3.31	3.40	3.30	3.50	3.43	3.48	3.42
Asian – Begin	2.53	2.51	2.53	2.53	2.54	2.55	2.54	2.53
Asian – End	3.02	3.13	2.98	3.10	2.85	2.93	2.85	2.93

Notes: Racial homophily % refers to the percent of all friendship ties in the model that are between racially-concordant individuals. Clustering % refers to the percent of all social ties where an ego and connected alter each share a tie with a third individual (i.e., % friend-of-a-friend). ‘Begin’ outcomes are calculated at model time step 7; ‘end’ outcomes are calculated at time step 497 (70 weeks later).

70 model weeks. At initialization, soda consumption is greatest among Latinos (~3.6 days per week) and lowest among Asians (~2.5 days per week). In Scenarios 1 through 4, which lack homophily preferences, soda consumption increases among Asians and whites (because their group-specific means start out lower than the mean of all agents in the model); consumption among Latinos drops by about 0.2 days per week. Despite these changes over time, social influence does not result in convergence between groups. This may be due to the relatively high level of segregation in the model, which results in a relatively high level of racial homophily even in the absence of preferences for homophily. In contrast, the group-specific changes in SSB consumption are smaller in magnitude in models with homophily preferences. For example, in Scenario 1 baseline SSB consumption among Asians 2.53 relative to 3.02 after 70 weeks had elapsed, a difference of 0.49 days/week. In Scenario 5, which included homophily preferences, the change was 0.31 days per week. This suggests that, given levels of segregation in the model, between-group inequalities are more persistent given more homophilous social networks.

Scenarios that included a family influence showed greater changes over time among Asians (who had lowest SSB consumption at baseline) and Latinos (highest). For example, the declines in SSB consumption were greater among Latinos in scenarios that included a family influence. This may be due to complementary effects of the dual forms of social influence.

Table 4 presents the results of an analogous set of four interventions as simulated in BV. The primary differences are that in the Philadelphia model there is family influence in all scenarios, intervention participants include Latino individuals who are frequent consumers of soda, and the ‘neighborhood’ intervention consists of a decrease in the mRFEI of predominantly-Latino tracts with unhealthy food environments at baseline.

Results from the Philadelphia model are similar to those for BV in that there is considerable recidivism to pre-intervention soda consumption levels during the course of model runs. In the Philadelphia model, this is true for all scenarios. In contrast to BV, there is little difference in durability of intervention effects by type of intervention. This may be due to the fact that family and peer influence are strong enough to consolidate behaviors before the intervention; as such, participants are strongly influenced by the behaviors of their friends and family after the intervention occurs.

#### 4. Discussion

In this study, I examined some of the implications that the interrelationships between residential segregation, social network

**Table 4**  
Social network characteristics and durability of effects of an intervention to reduce soda consumption among Latinos in Philadelphia.

	Scenario 1				Scenario 4			
<b>Model Parameters</b>								
Homophily Pref.	0				25			
Clustering Pref.	0				25			
Family Influence	Yes				Yes			
	Intervention Type				Intervention Type			
	R/NF	R/F	N/NF	N/F	R/NF	R/F	N/NF	N/F
<b>Weekly Soda Consumption</b>								
Pre-Intervention	3.51	3.55	3.42	3.52	3.58	3.61	3.53	3.58
Post-Intervention	0.91	0.92	0.94	0.90	0.93	0.96	0.95	0.95
End	3.19	3.21	3.14	3.14	3.24	3.14	3.00	3.08

Notes: Pre-intervention is at time step 490; post-intervention is at 511; end is at 1001. R=randomly-targeted intervention; N=network-based intervention; NF=no food environment intervention; F=food environment intervention.

formation, and social influence may have on the persistence of health inequalities. I used the stylized BV model to identify segregation and network characteristics that allow inequalities either to attenuate or persist over time, as well as to understand whether social influence may impact the durability of intervention effects. I used the Philadelphia model to investigate these same questions and frame the BV findings within the context of a real health behavior given levels of segregation and population characteristics that reflect those in an actual American city.

Results from the BV model suggest that, unsurprisingly, the combination of segregation and homophily are integral to the persistence of health inequalities. This finding is highly consistent with the health disparities literature (Williams & Collins, 2001; Williams & Jackson, 2005). Perhaps more surprising are findings from BV suggesting that, in the absence of residential segregation, social influence can reduce or eliminate even large inequalities produced by between-group differences in initial preferences. Inequalities were greatly reduced even in scenarios with preferences for homophily but not segregation, and nearly eliminated in those with preferences for neither segregation nor homophily. In contrast, inequalities largely persisted in scenarios with preferences for segregation, even absent preferences for homophily or clustering. This is likely the joint result of the fact that residential segregation can produce segregated social networks (i.e., homophily) in and of itself, even without actual preferences for homophily, as well as the fact that a consequence of segregated neighborhoods is between-group differences in access to resources at the neighborhood level.

The results from the Philadelphia model were illustrative because they highlighted the relative importance of homophily and clustering given plausible levels of segregation that might be observed in a large American city. As might be expected given the findings of BV, between-group inequalities in the Philadelphia model were much more durable over time. This is likely because the relatively high level of segregation in the Philadelphia model yielded fairly high levels of homophily in all scenarios. This was particularly true among the two largest groups, Whites and Blacks, who demonstrated almost no change in behaviors in every scenario. In contrast, changes over time were small but meaningful and consistent among Latinos and Asians, the two smallest groups. In scenarios that lacked preferences for homophily, members of these groups had the most heterogeneous social networks.

These models illustrate at least two general points: First, inequalities in behaviors subject to social influence seem most likely to persist over time when social networks are both homophilous and groups are segregated across neighborhoods. Second, residential segregation in and of itself appears to be enough to generate homophily sufficient for the persistence of inequalities. This finding may be important for understanding the long-term social pathology of inequalities in the U.S., particularly given the high levels of residential, school-based, occupational, and other forms of segregation in most American cities.

Others have used ABMs and other simulation frameworks to examine social influences on health behaviors, as well as mechanisms that may produce dietary and other health behavior inequalities (Auchincloss et al., 2011; Hammond & Ornstein, 2014; Orr et al., 2014; Shoham et al., 2012; Trogdon & Allaire, 2014; Yang et al., 2011; Zhang, Shoham, Tesdahl, & Gesell, 2015; Zhang et al., 2015). Shoham, Zhang and colleagues have used various simulation models to disentangle the respective roles of homophily, shared environment, and peer influence in creating health behavior ‘clusters’ within social networks, as well as to understand how social influence can be leveraged to develop more effective interventions (Shoham et al., 2012; Zhang, Shoham, Tesdahl, & Gesell, 2015; Zhang et al., 2015). Auchincloss and colleagues use a stylized ABM to understand how food environment disparities and initial differences in preferences between groups can interact to produce long-term inequalities in diet (Auchincloss et al., 2011).

Only one other study of which I am aware has examined how social

network formation and social influence may contribute to health inequalities (Orr et al., 2014). Orr and colleagues explored how education policy, social network effects, and perceptions of social norms interact to increase, decrease, or maintain health inequalities between black and white students. The focus of the study was primarily on understanding the extent to which positive or negative social norms (e.g., students' perceptions of whether the 'norm' healthy vs unhealthy behavior) impact inequalities. A key contrast between the Orr study and the current study is that I focus on understanding how varying levels of segregation and preferences for social network formation may contribute to inequalities. In the Orr study, the mechanisms driving segregation and social network formation were each static. In contrast, Orr used varying mechanisms of social influence (i.e., positive vs. negative norms), whereas in this study I use a static 'follow-the-average' model of social influence that more closely resembles that used by Hammond and Ornstein (2014) and Yang and colleagues (2011).

The findings of this study may have implications for health behavior interventions, particularly those seeking to reduce inequalities or that are targeting behaviors subject to social influence. Findings from both models suggest that even highly effective interventions targeting behaviors subject to social influence should expect a great deal of recidivism over time. This is particularly true of interventions targeted towards individuals embedded in social networks that exhibit unhealthy behaviors. Since historical circumstances, socioeconomic deprivation, and unhealthy environments have often interacted to produce unhealthy behaviors at the group level, this may be particularly important for health inequalities. This study also shows that positive changes in behavior can be made more durable via network-based recruiting into interventions, particularly when social networks are characterized by a high degree of homophily and clustering. Finally, the models provide support for multilevel interventions that include both an individual and neighborhood-level component.

This study has potentially important limitations. As with any simulation model, the validity of this study's conclusions is limited by the extent to which mechanisms in the model accurately represent those at work in the real world. Including every sub-process relevant to segregation, social network formation, and social influence in an ABM is probably not desirable, as increased complexity typically leads to difficulties in interpreting and understanding model results. As such, I have not included sub-processes like residential mobility that affect segregation patterns in cities (Bruch, 2014), formation or dissolution of friendships that affect social network topology (Shoham et al., 2012), and others. As a mechanistic study, I do not intend these models to accurately predict real-world levels of bubblegum chewing or soda consumption. Rather, I intend this work to highlight potential mechanisms through which segregation, social network formation, and social

influence may interact to contribute to inequalities. While the BV model is highly stylized in almost every regard, the Philadelphia model uses mechanisms of network formation and social influence that likely do not reflect the complexity of these phenomena in the real world. For example, the 'follow the average' mechanism I use in the models obscures the fact that people do not always view their friends' behaviors, that some friends' behaviors or attitudes may be more accurately viewed than others', and that some friends may be more influential than others. I use this simplified mechanism for ease of interpretation, because as the level of complexity in ABMs increases it becomes increasingly difficult to understand whether outcomes demonstrated by the model are the result of one component of the model, interrelationships between multiple components of the model, or an artifact of programming or arbitrary decisions made during the model development process. Furthermore, in the Philadelphia model I used data from a national study to generate a plausible distribution of SSB consumption at model initialization. A limitation of the Philadelphia model is that these national data may not accurately represent SSB consumption behavior in Philadelphia. In general, whether or not you believe that meaningful conclusions can be drawn from these models will likely reflect your belief that the models do or do not meaningfully capture important aspects of social interaction and behavior.

## 5. Conclusions

This study has implications for persistent health inequalities. The models suggest that in the absence of segregation, even relatively small levels of social network ties between racial/ethnic groups would produce reductions in inequality over time. An unsurprising but important conclusion is that residential segregation is a fundamental process that enables the production and persistence of health inequalities present in most large urban areas. The other major conclusion of the study is that, for behaviors that are subject to social influence and that cluster within social groups, interventions that do not include multiple network members will likely experience some (large) degree of recidivism. As discussed, this may be particularly salient for interventions targeting populations that are jointly influenced by poor social conditions and unhealthy environments. Taken together, these conclusions reinforce the idea that reducing or eliminating inequalities will require broad, multilevel, sustained intervention.

## Acknowledgements

This work was supported by the Robert Wood Johnson Foundation New Connections program (grant number 73267). The funder had no role in the design, analysis or writing of this study.

## Appendix A

### A.1. Bubblegum Village

Bubblegum Village (BV) is a stylized model that includes agent-individuals from two groups: a green majority and a red minority. I use BV to explore implications of extreme (i.e., very high or very low) preferences for segregation, homophily, and clustering. Each agent in BV lives in a neighborhood, has a social network of between three and five friends drawn from their neighborhood and adjacent neighborhoods, and makes a daily decision regarding whether or not to chew bubblegum. The model allows for varying levels of segregation preferences, homophily preferences (e.g., the preference of a green person to be friends with other green people), and clustering (i.e., increase in the proportion of 'friends of friends'). Upon initialization, population-level bubblegum preferences differ between the red and green groups and neighborhood-level barriers and facilitators differ between majority-green and majority-red neighborhoods.

BV is a metaphor for complex systems characterized by the following: (1) existence of majority and minority groups, (2) group-level differences in initial preferences, (3) neighborhood-level influence, (4) social influence. Examples of systems that generally meet these criteria might include diet, physical activity, or substance use behaviors in most American cities. The purpose of BV is to explore how residential segregation, neighborhood-level influences, social influence, and initial between-group differences in behavioral preferences interact to produce disparities

between majority and minority groups.

#### A.1.1. Agents and environment

BV is comprised of 2250 individuals spread across a 500×500 continuous space. Each individual is assigned a color: 70% are the green majority and 30% are the red minority. Each individual is assigned to a housing unit within one of 100 neighborhoods. Housing units are uniformly distributed across space, so neighborhoods each contain the same number of units.

#### A.1.2. Segregation mechanism

Individuals select from available housing units with probability of selection proportionate to the utility associated with that unit. Utility for a housing unit is based on three factors: (1) whether the majority of occupants in the destination neighborhood are color-concordant with the individual making the selection (e.g., green people prefer to live in majority-green neighborhoods), (2) whether surrounding neighborhoods are color-concordant with the individual, and (3) a population-level segregation preference that represents the extent to which agents in the model prefer to live in a neighborhood with agents of the same color. Segregation preference varies between model runs, but I use a value of zero for iterations of the model with no preference for segregation and a value of 10 for iterations with a high preference for segregation. The utility ( $U_{ij}$ ) that agent  $i$  associates with a housing unit in neighborhood  $j$  is given by

$$U_{ij}=1 + S(C_{ij}) + \sum_1^w \frac{S(N_{ik})}{w}$$

where  $S$  is the segregation preference,  $C_{ij}$  is a dummy variable representing color concordance between agent  $i$  and neighborhood  $j$ ,  $w$  is the number of neighborhoods adjacent to  $j$  (typically four), and  $N_{ik}$  is a dummy variable representing color concordance between agent  $i$  and adjacent neighborhood  $k$ . Once utilities are calculated for a set of potential destination housing units, a single unit is selected from the set with probability of selection proportionate to each unit's utility.

As an example, if  $S$  has a value of 10, the value of  $U_{ij}$  can range from a value of one for a housing unit in a neighborhood that is not color-concordant and no color-concordant adjacent neighborhood, to a value of 21 for a housing unit in a neighborhood that is color concordant with adjacent neighborhoods that are all color concordant. The purpose of this utility model is to facilitate a neighborhood selection process through which the value of  $S$  can be manipulated to produce an environment with no segregation (i.e., random), moderate segregation, and relatively high levels of segregation. Also note that including an increase in utility associated with color concordance of adjacent neighborhoods produces a smoothing in the color distribution of neighborhoods across neighborhoods.

#### A.1.3. Social network formation

Each agent in the model has a number of close social contacts (i.e., friends) that comprise their social network. All friendships are bi-directional. Previous work from the General Social Survey suggests that American adults have an average of 2.1 close social contacts, while a panel study conducted by O'Malley and colleagues (2012) suggests a mean of 4.4 close social contacts (McPherson, Smith-Lovin, & Brashears, 2006; O'Malley, Arbesman, Steiger, Fowler, & Christakis, 2012). In the model, I use a random draw from a uniform distribution to assign each individual a 'target' number of friends between three and five.

Friends are chosen from an array of potential friends that includes (1) people who live in the same neighborhood, (2) people who live in an adjacent neighborhood, and (3) friends of existing friends. All potential friends are assigned a 'friendship utility' based on (1) homophily, or color concordance between the two agents (i.e., red or green) and (2) clustering, or whether two agents share a friend in common. The increase in utility associated with each factor is based on population-level preferences for homophily. The utility ( $U_{ij}$ ) of a potential connection between agent  $i$  and agent  $j$  is given by

$$U_{ij}=1 + H(Conc_{ij}) + C(F_{ij})$$

where  $H$  is homophily preference,  $Conc_{ij}$  is a dummy variable representing color concordance between agents  $i$  and  $j$ ,  $C$  is a clustering preference, and  $F_{ij}$  is a dummy variable representing whether agents  $i$  and  $j$  share a friend in common. Once utilities are derived for the entire set of potential friends, one friend is selected with probability proportionate to friendship utility.

#### A.1.4. Bubblegum chewing behavior

The probability ( $p_i$ ) that individual  $i$  will chew bubblegum on a given day is a function of an individual-level attitude ( $A_{ij}$ ) and a neighborhood-level influence ( $N_j$ ) that reflects barriers or facilitators to chewing:

$$p_i=A_{ij}+N_j$$

To assign individual-level attitudes at model initialization, individuals are designated as 'frequent chewers' or 'infrequent chewers' at model initialization. Attitudes of frequent chewers are drawn from a uniform distribution: U(0.8, 0.95). Attitudes of infrequent chewers are drawn from: U(0.05,0.15). For all model runs, 75% of red individuals are designated as frequent chewers and 25% as infrequent chewers; among greens the reverse is true.

Each neighborhood in the model is defined at initialization as a barrier or facilitator of bubblegum chewing behavior. Facilitator neighborhoods increase the probability that residents will chew bubblegum, while barrier neighborhoods decrease this probability. Among facilitators, neighborhood influence is drawn from U(0.05, 0.15); among barrier neighborhoods, influence is drawn from U(-0.15, -0.05). For all model runs, 75% of red-majority neighborhoods are facilitators and 25% barriers; in green-majority neighborhoods, 25% are facilitators and 75% barriers.

The value of  $p_i$  is constrained in the range [0.01,0.99] so that all agents have a non-zero probability of chewing bubblegum but are not certain to chew bubblegum.

#### A.1.5. Social influence

The influence of social network members on bubblegum chewing behavior operate via feedback loops. These feedback loops, adapted from Yang (2011), update each individual's attitudes ( $A_{ij}$ ) based on the behaviors of network members. Once per week (i.e., every seven model time steps), a



small proportion ( $a=0.05$ ) of an individual's attitudes are updated based on the mean observed behaviors of network members:

$$A_{ij}' = (1 - a) \times A_{ij} + a \overline{Y_f(F)}$$

where  $Y_f$  represents the proportion of days in the previous week during which social network member  $f$  chewed bubblegum. Because behavioral outcomes are governed by both attitudes and neighborhood factors, the feedback loops may have different effects under different model conditions.

#### A.1.6. Intervention scenarios

BV includes three intervention scenarios to reduce bubblegum chewing among the red minority group and thus decrease between-group inequalities. Intervention 1 improves attitudes ( $A_{ij}$ ) among 99 randomly-selected members of the red group. The model first tries to target 'bad behaviors' with  $A_{ij} > 0.75$  at  $T_0$ ; if there are fewer than 99 'bad behaviors' the model then selects randomly-selected members of the red population. Intervention 2 is a 'network-based' intervention that improves ( $A_{ij}$ ) among 33 primary participants drawn from the pool of 'bad' behaviors from the red group, as well as two members of each primary participant's social network (total  $n=99$ ). In Interventions 1 and 2, the ( $A_{ij}$ ) of intervention participants is 'improved' to a value determined via a random draw from  $U(0.05, 0.15)$ . Intervention 3 is an 'environmental' intervention that addresses neighborhood-level facilitators of bubblegum chewing. Specifically, the intervention reduces the influence of 'facilitator' neighborhoods by an amount drawn from a uniform distribution:  $U(0.15, 0.05)$ . Since bubblegum chewing is a function of neighborhood influence, this effectively reduces bubblegum-chewing probabilities of all people in 'bad' neighborhoods, should have a greater impact on the red population than the green population, and should thus reduce inequalities. In addition to these three intervention scenarios, I also test 'multilevel' interventions that simultaneously include the individual-level interventions and the food environment intervention. All interventions are initiated after 500 time steps to allow for a 'warm up' period in which individual preferences reach stability (i.e., individual preferences are no longer changing due to social influence). I determined this warm up period via visual inspection of a running means plot of color-group specific bubblegum chewing during the baseline scenario.

#### A.1.7. Model iterations

To understand how different combinations of segregation and social network formation impact the persistence of disparities, I use a  $2 \times 2 \times 2$  factorial design that includes strong vs. no segregation preferences, strong vs. no preferences for color homophily (e.g., reds want to be friends with reds), and strong vs. no preferences for clustering (i.e., friends of friends). To examine the durability of intervention effects, I use a  $3 \times 2$  design that includes individual-level interventions (no intervention, random intervention, network-based intervention) and a food environment intervention. I run these interventions in two environments: one with no preferences for segregation or network formation, and a second with high preferences for segregation, clustering, and homophily.

#### A.1.8. Model outcomes

Each individual makes a daily (i.e., every model time step) decision about whether or not to chew bubblegum. To assess differences across different iteration of the model, I collect mean weekly (i.e., every seven time steps) bubblegum consumption for the red and green populations. I average these outcomes across 25 replications (i.e., a single 'run' of the model) for each iteration of the model.

### A.2. ABM of sugar-sweetened beverage (SSB) consumption in Philadelphia

In addition to the stylized BV model, I also present results from an ABM of SSB consumption in Philadelphia, PA. The purpose of the Philadelphia model is to examine whether social network characteristics have implications for production and persistence of health disparities given observed levels of segregation in an American city and plausible between-group differences in an important health behavior. The model is based in GIS space, with a population of 19,319 individuals (i.e., 2% of the Philadelphia population) distributed across Philadelphia census tracts.

#### A.2.1. Agents and environment

The primary agents in the Philadelphia ABM are individuals living in households. The population of individuals is synthetic, meaning that no agent corresponds directly with a person in the real world; however, the distribution of the population and assignment of socio-demographic characteristics is based on observed data. Specifically, household composition (i.e., family vs. non-family households, number of adults, number of children) and individual characteristics are assigned proportionate to census tract-level distributions observed in the 2010–2014 American Community Survey and the 2010 Decennial Census. Each individual is assigned age, gender, marital status, educational attainment, income, and race/ethnicity. Race/ethnicity is restricted to the four largest groups in Philadelphia: white, black, Latino, and Asian. The model includes children and adults. Children are assigned to the nearest elementary/middle or high school, depending upon their age, and adults are assigned to a random workplace. The location of schools is determined based on actual school locations in Philadelphia (as determined by data from the National Center for Education Statistics); workplaces are randomly distributed throughout the city. The model also includes a composite measure of the food environment at the census tract level, the modified Retail Food Environment Index, derived from Babey, Wolstein & Diamant (2011). Data on food outlets in Philadelphia census tracts are from Dun and Bradstreet (2012), a commercial business listing service, from 2012.

#### A.2.2. Diet outcomes

At model initialization, each agent is assigned a baseline probability,  $p_i$  that they will consume at least one SSB on a given day. The purpose of assigning this probability is so that at baseline, the population of agents in the model have a plausible distribution of SSB consumption given observed data.  $p_i$  is a function of agents' individual, household, and neighborhood characteristics, as well as a weight of the importance of each characteristic for SSB consumption:

$$p_i = I_i b_I + H_i b_H + E_i b_E + b_0 + \varepsilon$$

where  $I_i$  is individual characteristics (e.g., age, gender, educational attainment, marital status, race/ethnicity),  $H_i$  is household characteristics (e.g., income),  $E_i$  is environmental factors related to access to food resources,  $b_0$  is an intercept, other  $b$ s are weights of the importance of each factor for SSB consumption, and  $\varepsilon$  is a random component that represents heterogeneity in individuals on unobserved characteristics (e.g., attitudes and beliefs) that impact diet outcomes.

**Table A.1**

Logistic regression parameters derived from the National Health and Nutrition Examination Survey used for predicting baseline sugar sweetened beverage consumption among adults and children in the Philadelphia model.

	<b>b. Adults</b>	<b>b. Children</b>	<b>Source</b>
<b>Age (yr)</b>	-0.0347	0.0701	
<b>Male</b>	0.538	0.433	
<b>Marital Status</b>			
Married	Ref.	–	
Never Married	-0.294	–	
Div/Wid/Sep	0.0907	–	
<b>Education</b>			
< HS	Ref.	–	
High School	-0.0677	–	
College Grad.	-0.737	–	
<b>Income (% FPL)</b>	-0.144	-0.152	
<b>Race/Ethnicity</b>			
White	Ref.	Ref.	
Black	0.138	-0.664	
Latino	0.123	-0.121	
<b>Model Intercept (<math>b_0</math>)</b>	1.367	0.555	

Note: FPL=Federal Poverty Level. Please note that parameter values are based on a logistic regression model.

I use a logistic regression model applied to data from the 2007 to 2010 National Health and Nutrition Examination Survey (NHANES) to estimate weights attributed to each characteristic (i.e., the  $b$ 's in the above equation). Note that I estimate these weights separately for adults and children. Table A.1 includes a list of values derived from the logistic regression model. Similar to the BV model, the value of  $p_i$  is constrained in the range [0.01,0.99] so that all agents have a non-zero probability of SSB consumption on a given day, but are not certain to consume an SSB.

I was unable to estimate the parameter representing the impact of the food environment on soda consumption ( $b_E$ ) directly using NHANES. I thus use values in line with estimates from the Babey et al. (2011). Specifically, agents' probability of consuming SSB on a given day increases by one percentage point for each one-point increase in the modified Retail Food Environment Index (Table A.2).

**Table A.2**

Key parameters in the Bubblegum Village and Philadelphia models.

Parameter	Value
<i>Bubblegum Village</i>	
Population size	2250
Percent minority	30
Social network size	U(3, 5)
Segregation preference ( $S$ )	Varies across iterations: 0 or 10
Homophily preference ( $H$ )	Varies across iterations: 0 or 10
Clustering preference ( $C$ )	Varies across iterations: 0 or 10
Daily probability of chewing bubblegum ( $p_i$ )	Sum of $A_{ij}$ and $N_j$ ; constrained to [0.01, 0.99]
Attitudes towards bubblegum chewing at baseline ( $A_{ij}$ )	Frequent chewers: U(0.8, 0.95) Infrequent chewers: U(0.05,0.15)
Proportion frequent chewers	Red: 75% Green: 25%
Neighborhood influence ( $N_j$ )	–
Proportion facilitator neighborhoods	Majority-red neighborhoods: 75% Majority-green neighborhoods: 25%
Social influence magnitude ( $a$ )	0.05
<i>Philadelphia Model</i>	
Population size	19,319 (2% of Philadelphia population)
Social network size	U(3, 5)
Homophily preference ( $H$ )	Varies across iterations: 0 or 25
Clustering preference ( $C$ )	Varies across iterations: 0 or 25
Daily probability of SSB consumption at baseline ( $\hat{y}_i$ )	Constrained to (0.01, 0.99); baseline value derived based on weights (see Table A.1)
Friend social influence ( $a_{fr}$ )	0.05
Family social influence ( $a_{fa}$ )	Varies across iterations: 0 or 0.05

### A.2.3. Social network formation

In the Philadelphia model, individuals each have a ‘friend’ and ‘family/household’ network. Family networks consist of everyone that lives in the same housing unit. Friendship selection is similar to the BV model. Main differences include that the array of potential friends includes people who live in the same census tract (i.e., neighbors), friends of friends, and schoolmates for children or co-workers for adults. Friends are chosen based on a ‘friendship utility’ that includes a fixed preference for similar age and gender, as well as a variable preference for racial homophily and clustering that ranges from a value of 0 (no preference) to a value of 25 (high preference). Individuals in the model are assigned friends at model initialization until they reach an a target number of friends based on a random draw from a uniform distribution between three and five.

### A.2.4. Social influence

Social influence of family members and peers operates largely the same in the Philadelphia model as in BV. The only exception is that, in the Philadelphia model, an individual's preferences can be updated based on the average preferences of the friend network and the average preferences of the family network:

$$Y_i' = (1 - a_{fr} - a_{fa}) \times Y_i + a_{fr} \overline{Y_{fr}(Fr)} + a_{fa} \overline{Y_{fa}(Fa)}$$

In the current study I use  $a_{fa}=0$  and  $a_{fr}=0.05$  (i.e., friend influence but no family influence) for some iterations of the model for all simulation runs and  $a_{fa}=0$  and  $a_{fr}=0.05$  for others (i.e., equal friend and family influence).

### A.2.5. Intervention scenarios

Intervention scenarios in the Philadelphia model are similar to BV, but pertain only to SSB consumption. Intervention reduces the probability that each participant will consume SSB by a factor with uniform distribution:  $U(0.15, 0.05)$ . The ‘random’ intervention includes 99 individuals randomly selected from the sub-population of Latinos with high probability of consuming SSB. The network-based intervention includes 33 Latino frequent SSB consumers, with two additional friends or family members per primary participant. For the food environment intervention, I improve the overall food environment in Latino neighborhoods. Specifically, in all census tracts with baseline mRFEI  $\geq 0.5$  and  $\geq 40\%$  Latino residents, I subtract  $U(1, 3)$  from the baseline mRFEI. The baseline range of mRFEI in Philadelphia census tracts is between 0 and 100; I constrain the post-intervention value to be a minimum of 0. All interventions are initiated after 500 time steps to allow for a ‘warm up’ period. I determined this warm up period via visual inspection of a running means plot of race/ethnicity specific bubblegum chewing during the baseline scenario.

### A.2.6. Model iterations

I use a  $2 \times 2$  design that includes strong vs. no racial homophily preference and strong vs. no clustering preference to examine the persistence of SSB consumption inequalities between racial/ethnic groups. All model runs have fixed levels of residential segregation and distributions of individual and household characteristics. Given this constraint, I explore how social influence under different mechanisms of social network formation may contribute to persistence of disparities. The primary outcome is the mean number of days in the previous week with at least one SSB consumed, stratified by race/ethnicity. I assess outcomes across an average of 20 simulation runs.

## A.3. Key assumptions in the BV and Philadelphia models

I make several key assumptions in both the BV and Philadelphia models that are important to consider. In general, a potential strength of simulation models is that the coding process forces the modeler to explicitly consider and make decisions regarding assumptions as the model is implemented. One assumption is that all agents in a given model have equal preferences related to segregation, homophily, and clustering. While this is clearly implausible, the purpose of the segregation and social network formation processes is to create neighborhoods and social networks with higher or lower levels of segregation, homophily, and clustering. While the preferences are implausibly uniform, the resulting neighborhoods and networks are not. Since preferences only affect these initial conditions, so the uniformity of preferences has no impact beyond initial model conditions. I also assign agents a baseline probability of either chewing bubblegum or consuming SSB on a given day. In BV the assignment is based on agents’ colors, while in the Philadelphia model I assign probabilities based on agents’ individual and neighborhood characteristics and weights derived from a cross-sectional study of diet. In both cases, I make an assumption that agents’ characteristics impact the baseline probabilities, but that the only mechanism of ongoing influence over time (as the model runs) operates via social influence. A further assumption is that social influence between individuals occurs via a ‘follow the average’ mechanisms whereby an agent's behavior is influenced by the average behaviors of friends and family. Similarly, I allow all agents to have perfect information regarding the behaviors of their friends and family. I exclude other potential sources (e.g., more diffuse peer groups, media) and forms (e.g., satisficing mechanism presented in [Hammond and Ornstein \(2014\)](#)) of social influence. I also assume that individuals have underlying preferences for housing (BV model) and friendship formation (both models) that actually impact the choices they make. In the case of friendship formation, I assume that agents: (a) have a target number of friends, (b) select between friends based on their underlying preferences, and (c) that the pool of ‘potential’ friends is restricted based on geography (BV model) or geography, workplace, and school (Philadelphia model).

## A.4. Sensitivity analyses

I conducted a range of sensitivity analyses to help identify the extent to which results of the models reflect initial model conditions or sensitive parameters. I ran a series of parameter variation experiments in both models to determine the extent to which the size of social networks may impact results. Specifically, I conducted multiple runs of model iterations with variation in the parameters governing the lower (range=0 to 3) and upper (range=5 to 8) limits on the size of friendship networks. To assess sensitivity to this parameter, I examined changes in group-specific mean outcomes after 490 model time steps (i.e., 70 weeks) across 5 replications of each iteration. These analyses suggest that the model results are highly insensitive to the size of friendship networks, likely due to the ‘follow the average’ mechanism of social influence used in the model.

I also conducted a series of experiments to understand whether the substantive implications of BV as impacted by the initial distribution of neighborhoods (e.g., 75% of red neighborhoods were ‘facilitators’ and 25% ‘barriers’ to bubblegum chewing) and agent preferences. Specifically, I examined the same 8 scenarios summarized in [Table 1](#) but included a fixed proportion of ‘moderate’ bubblegum chewers ( $A_j$  distributed uniformly

**Table A.3**

Baseline conditions for additional Bubblegum Village experiments that include moderate chewers and neutral neighborhoods at initialization.

	Experiments in Table 1	Additional Experiments
<b>Neighborhood Influence</b>		
Red	25% barriers; 75% facilitators	25% barriers; 50% facilitators; 25% neutral
Green	75% barriers; 25% facilitators	50% barriers; 25% facilitators; 25% neutral
<b>Initial Preferences of Agents</b>		
Red	25% infrequent chewers; 75% frequent chewers	25% infrequent chewers; 50% frequent chewers; 25% moderate chewers (range of)
Green	75% infrequent chewers; 20% frequent chewers	50% infrequent chewers; 25% frequent chewers; 25% moderate chewers

from 0.45 to 0.55) and ‘neutral’ neighborhoods ( $N_i$  distributed uniformly from  $-0.05$  to  $0.05$ ). The baseline conditions of these models are summarized in Table A.3. Effectively, the primary difference between results in these additional experiments and those reported in the manuscript is that between-group differences were much smaller due to the inclusion of moderate chewers and neutral neighborhoods. However, the main findings with respect to persistence of between-group inequalities and persistence of intervention effects were largely the same, irrespective of these differences in initial model conditions. As such, I chose to include experiments based on the more ‘extreme’ scenarios in the manuscript for illustrative purposes.

## References

- Auchincloss, A. H., Riolo, R. L., Brown, D. G., Cook, J., & Diez Roux, A. V. (2011). An agent-based model of income inequalities in diet in the context of residential segregation. *American Journal of Preventive Medicine*, *40*(3), 303–311.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall, Inc..
- Babey, S. H., Wolstein, J., & Diamant, A. L. (2011). *Food Environments Near Home and School Related to Consumption of Soda and Fast Food* Los Angeles, CA: UCLA Center for Health Policy Research.
- Batis, C., Hernandez-Barrera, L., Barquera, S., Rivera, J. A., & Popkin, B. M. (2011). Food acculturation drives dietary differences among Mexicans, Mexican Americans, and Non-Hispanic Whites. *The Journal of Nutrition*, *141*(10), 1898–1906.
- Boone-Heinonen, J., Gordon-Larsen, P., Kiefe, C. I., Shikany, J. M., Lewis, C. E., & Popkin, B. M. (2011). Fast food restaurants and food stores longitudinal associations with diet in young to middle-aged adults: The CARDIA study. *Archives of Internal Medicine*, *171*(13), 1162–1170.
- Bowman, S. A., Gortmaker, S. L., Ebbeling, C. B., Pereira, M. A., & Ludwig, D. S. (2004). Effects of fast-food consumption on energy intake and diet quality among children in a national household survey. *Pediatrics*, *113*(1), 112–118. <http://dx.doi.org/10.1542/peds.113.1.112>.
- Bruch, E. E. (2014). How population structure shapes neighborhood segregation. *American Journal of Sociology*, *119*(5), 1221.
- Card, D., & Rothstein, J. (2007). Racial segregation and the black–white test score gap. *Journal of Public Economics*, *91*(11–12), 2158–2184. <http://dx.doi.org/10.1016/j.jpubeco.2007.03.006>.
- Christakis, N. A., & Fowler, J. H. (2007). The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, *357*(4), 370–379.
- Christakis, N. A., & Fowler, J. H. (2008). The collective dynamics of smoking in a large social network. *New England Journal of Medicine*, *358*(21), 2249–2258.
- Cunningham, S. A., Vaquera, E., Maturu, C. C., & Narayan, K. V. (2012). Is there evidence that friends influence body weight? A systematic review of empirical research. *Social Science Medicine*, *75*(7), 1175–1183.
- de Vet, E., de Ridder, D. T. D., & de Wit, J. B. F. (2011). Environmental correlates of physical activity and dietary behaviours among young people: A systematic review of reviews. *Obesity Reviews*, *12*(501), e130–e142. <http://dx.doi.org/10.1111/j.1467-789X.2010.00784.x>.
- Diez-Roux, A. V., Nieto, F. J., Caulfield, L., Tyroler, H. A., Watson, R. L., & Szklo, M. (1999). Neighbourhood differences in diet: The Atherosclerosis Risk in Communities (ARIC) Study. *Journal of Epidemiology and Community Health*, *53*(1), 55–63.
- Dun & Bradstreet. (2012). *Duns Marketing Information* Waltham, MA: Dun Bradstreet.
- Flegal, K. M., Carroll, M. D., Ogden, C. L., & Curtin, L. R. (2010). Prevalence and trends in obesity among U.S. adults, 1999–2008. *Journal of American Medical Association*, *303*(3), 235–241. <http://dx.doi.org/10.1001/jama.2009.2014>.
- Fowler, J. H., & Christakis, N. A. (2008). Dynamic spread of happiness in a large social network: Longitudinal analysis over 20 years in the Framingham Heart Study. *British Medical Journal*, *337*, a2338.
- Glanz, K., Sallis, J. F., Saelens, B. E., & Frank, L. D. (2005). Healthy nutrition environments: Concepts and measures. *American Journal of Health Promotion*, *19*(5), 330–333 (ii).
- Go, A. S., Mozaffarian, D., Roger, V. L., Benjamin, E. J., Berry, J. D., Borden, W. B., & Turner, M. B. (2013). Heart disease and stroke statistics—2013 Update: A report from the American Heart Association. *Circulation*, *127*(1), e6–e245. <http://dx.doi.org/10.1161/CIR.0b013e31828124ad>.
- Hammond, R. A., & Ornstein, J. T. (2014). A model of social influence on body mass index. *Annals of the New York Academy of Sciences*, *1331*(1), 34–42.
- Holsten, J. E. (2009). Obesity and the community food environment: A systematic review. *Public Health Nutrition*, *12*(3), 397–405. <http://dx.doi.org/10.1017/S1368980008002267>.
- Larson, N. I., Story, M. T., & Nelson, M. C. (2009). Neighborhood environments: Disparities in access to healthy foods in the U.S. *American Journal of Preventive Medicine*, *36*(1), 74–81. <http://dx.doi.org/10.1016/j.amepre.2008.09.025>.
- Link, B. G., & Phelan, J. (1995). Social conditions as fundamental causes of disease. *Journal of Health and Social Behavior*, 80–94.
- Mabry, P. L., Marcus, S. E., Clark, P. I., Leischow, S. J., & Méndez, D. (2010). Systems science: A revolution in public health policy research. *American Journal of Public Health*, *100*(7), 1161–1163.
- Massey, D. S., Condran, G. A., & Denton, N. A. (1987). The effect of residential segregation on black social and economic well-being. *Social Forces*, *66*(1), 29–56.
- Massey, D. S., & Denton, N. A. (1988). The dimensions of residential segregation. *Social Forces*, *67*(2), 281–315.
- Massey, D. S., & Denton, N. A. (1989). Hypersegregation in US metropolitan areas: Black and Hispanic segregation along five dimensions. *Demography*, *26*(3), 373–391.
- Massey, D. S., & Denton, N. A. (1993). *American apartheid: Segregation and the making of the underclass*. Harvard University Press.
- Massey, D. S., & Fischer, M. J. (2000). How segregation concentrates poverty. *Ethnic and Racial Studies*, *23*(4), 670–691. <http://dx.doi.org/10.1080/01419870050033676>.
- McPherson, M., Smith-Lovin, L., & Brashears, M. E. (2006). Social Isolation in America: Changes in Core Discussion Networks over Two Decades. *American Sociological Review*, *71*, 353–375.
- Moore, L. V., Diez Roux, A. V., Nettleton, J. A., Jacobs, D. R., & Franco, M. (2009). Fast-food consumption, diet quality, and neighborhood exposure to fast food: The multi-ethnic study of atherosclerosis. *American Journal of Epidemiology*, *170*(1), 29–36. <http://dx.doi.org/10.1093/aje/kwp090>.
- Morland, K., Wing, S., Diez Roux, A., & Poole, C. (2002). Neighborhood characteristics associated with the location of food stores and food service places. *American Journal of Preventive Medicine*, *22*(1), 23–29.
- O’Malley, A. J., Arbesman, S., Steiger, D. M., Fowler, J. H., & Christakis, N. A. (2012). Egocentric social network structure, health, and pro-social behaviors in a national panel study of Americans. *PLoS One*, *7*, e36250.
- Orr, M. G., Galea, S., Riddle, M., & Kaplan, G. A. (2014). Reducing racial disparities in obesity: Simulating the effects of improved education and social network influence on diet behavior. *Annals of Epidemiology*, *24*(8), 563–569.
- Rosenquist, J. N., Murabito, J., Fowler, J. H., & Christakis, N. A. (2010). The spread of alcohol consumption behavior in a large social network. *Annals of Internal Medicine*, *152*(7), 426–433.

- Shoham, D. A., Tong, L., Lamberson, P. J., Auchincloss, A. H., Zhang, J., Dugas, L., & Luke, A. (2012). An actor-based model of social network influence on adolescent body size, screen time, and playing sports. *PLoS One*, *7*(6), e39795.
- Trogdon, J. G., & Allaire, B. T. (2014). The effect of friend selection on social influences in obesity. *Economics Human Biology*, *15*, 153–164. <http://dx.doi.org/10.1016/j.ehb.2014.08.004>.
- Widener, M. J., Metcalf, S. S., & Bar-Yam, Y. (2013). Agent-based modeling of policies to improve urban food access for low-income populations. *Applied Geography*, *40*(0), 1–10. <http://dx.doi.org/10.1016/j.apgeog.2013.01.003>.
- Williams, D., & Collins, C. (2001). Racial residential segregation: A fundamental cause of racial disparities in health. *Public Health Reports*, *116*(5), 404.
- Williams, D., & Jackson, P. (2005). Social sources of racial disparities in health. *Health Affairs*, *24*(2), 325–334. <http://dx.doi.org/10.1377/hlthaff.24.2.325>.
- Wing, R. R., & Phelan, S. (2005). Long-term weight loss maintenance. *The American Journal of Clinical Nutrition*, *82*(1), 222S–225S.
- Wing, R. R., Tate, D. F., Gorin, A. A., Raynor, H. A., & Fava, J. L. (2006). A self-regulation program for maintenance of weight loss. *New England Journal of Medicine*, *355*(15), 1563–1571. <http://dx.doi.org/10.1056/NEJMoa061883>.
- Yang, Y., Roux, A. V. D., Auchincloss, A. H., Rodriguez, D. A., & Brown, D. G. (2011). A spatial agent-based model for the simulation of adults' daily walking within a city. *American Journal of Preventive Medicine*, *40*(3), 353–361.
- Zhang, D., Giabbanelli, P. J., Arah, O. A., & Zimmerman, F. J. (2014). Impact of different policies on unhealthy dietary behaviors in an urban adult population: An agent-based simulation model. *American Journal of Public Health*, e1–e6. <http://dx.doi.org/10.2105/AJPH.2014.301934>.
- Zhang, J., Shoham, D. A., Tesdahl, E., & Gesell, S. B. (2015). Network interventions on physical activity in an afterschool program: An agent-based social network study. *American Journal of Public Health*, *105*(S2), S236–S243.
- Zhang, J., Tong, L., Lamberson, P., Durazo-Arvizu, R., Luke, A., & Shoham, D. (2015). Leveraging social influence to address overweight and obesity using agent-based models: The role of adolescent social networks. *Social Science Medicine*, *125*, 203–213.