Journal of Computational Social Science
https://doi.org/10.1007/s42001-022-00183-7

RESEARCH ARTICLE

Dynamic simulation of social media challenge participation to examine intervention strategies

Amro Khasawneh[1] · Kapil Chalil Madathil[2,3] · Kevin M. Taaffe[3] · Heidi Zinzow[4] · Amal Ponathil[5] · Sreenath Chalil Madathil[6] · Siddhartha Nambiar[7] · Gaurav Nanda[8] · Patrick J. Rosopa[4]

Received: 16 May 2022 / Accepted: 27 August 2022

Abstract
Recently, the use of social media by adolescents and young adults has significantly increased. While this new landscape of cyberspace offers young Internet users many benefits, it also exposes them to numerous risks. One such phenomenon receiving limited research attention is the advent and propagation of viral social media challenges. Several of these challenges entail self-harming behavior, which combined with their viral nature, poses physical and psychological risks for the participants and the viewers. In this paper, we show how agent-based modeling (ABM) can be used to investigate the effect of educational intervention programs to reduce participation in social media challenges at multiple levels—family, school, and community. In addition, we show how the effect of these education-based interventions can be compared to social media-based policy interventions. Our model takes into account the "word of mouth" effect of these interventions which could either decrease participation in social media challenge further than expected or unintentionally cause others to participate. We suggest that educational interventions at combined family and school levels are the most efficient type of long-term intervention, since they target the root of the problem, while social media-based policies act as a retrospective solution.

Keywords Agent-based model · Web-based challenges · Self-injurious behavior · Behavior · Integrated behavioral model · Social media

✉ Amro Khasawneh
Khasawneh_an@mercer.edu

Extended author information available on the last page of the article

Published online: 16 September 2022

Introduction and background

After remaining stable or declining for nearly 2 decades, rates of depressive symptoms, suicide, and suicide-related outcomes for adolescents in the United States have started to dramatically increase [1, 2]. As of 2019, suicide was the second leading cause of death among teens and young adults in the US [3]. The latest evidence suggests that one causative factor behind this change is the concurrent increase in the use of social media [2]. A growing body of literature [4–7] has shown that social media can influence self-harming and suicidal behavior, posing a significant threat to public health and safety [8].

Social media challenges

One understudied phenomenon is the advent and propagation of viral social media challenges. There is a wide variety of challenges that have propagated across several social media platforms. These challenges, which vary in their inherent level of self-harm behavior and risk to individuals, include the Tide Pod Challenge (TPC), the Cinnamon Challenge (CC), the Amyotrophic Lateral Sclerosis (ALS) Ice Bucket Challenge (IBC), and the Blue Whale Challenge (BWC) [9–13].

Some of these challenges promote philanthropic causes and tend to present lower risks of harm for participants than others. For example, the risks associated with the ALS IBC, which raises awareness of and funding for finding a cure for ALS research, are considered minimal [13–17]. On the other hand, most social media challenges yield few positive results, yet may involve harm to youths and young adults. For example, adolescents have swallowed Tide Pods (chemical detergent pods used for laundry), a challenge that resulted in more than eight cases of poisoning as reported by the American Association of Poison Control Centers (AAPCC) [18]. The Cinnamon Challenge (CC) entails participants eating a spoon of cinnamon without drinking water for more than a minute. The problem is that cinnamon does not dissolve nor biodegrade in the lungs, as evidenced by studies of lab rats which experienced symptoms ranging from mild multifocal granulomatous inflammation to alveolar lipoproteinosis and alveolar cell hyperplasia [9, 19, 20]. For humans, the consequences are just as serious, because swallowing a large amount of cinnamon can cause pulmonary inflammation, allergic and irritant reactions, and in even more serious situations, hypersensitivity-induced asthma attacks, which can be fatal [9]. However, none of these potentially fatal consequences have stopped adolescents and young adults from participating in CC. As of 2013, there are more than 51,100 public YouTube clips of someone accepting this challenge, with some videos garnering more than 19 million views globally [9, 12]. Given the significant amount of controversy concerning these online challenges, there is little research on modeling these factors and studying the impact of targeted interventions to address this harmful challenge.

Interventions

Interventions are needed to reduce the spread of these viral challenges [12, 13, 15–17, 20, 21]. These interventions should use a multifaceted approach, rather than a single strategy to minimize unwanted adolescent behavior [12, 16, 17, 22]. To identify appropriate areas for intervention, we adopted the Intervention Wheel from the public health literature [21, 23, 24]. The interventions are grouped together based on similarities and the possibility of simultaneous use. The Intervention Wheel works as a starting point for researchers to consider a wide variety of possible interventions and identify the most appropriate one. Furthermore, the Intervention Wheel is supported by evidence, as it is verified by sound science and effective practices [23]. Each intervention on the wheel is clearly defined in a manner that describes the actions directly. Health teaching, for example, is a type of intervention that involves creating material to educate a specific population about the targeted behavior either on the system level, community level, or individual level.

Health teaching interventions are defined as "sharing information and experiences through educational activities designed to improve health knowledge, attitudes, behaviors (norms), and skills" [25, 26]. Specific interventions need to be driven by the factors influencing participation in harmful challenges and specific items that make up each of these factors such as attitudes and perceived norms. In a previous study, we found that adolescents view the activities in the challenges as enjoyable, value the consequences of such activities, and, due to the viral nature of these challenges, perceive participation in these activities as a way to receive approval from peers [12]. In other words, adolescents fail to foresee the negative consequences of these activities and instead overemphasize and overestimate the peer approval to be gained from participation in these challenges. Thus, the interventions we recommend address these correlates of challenge participation based on prior research [12]. In this paper, we provide an overview of the interventions. Details of the interventions are beyond the scope of this paper. In this work, we investigate the effect of deploying these interventions using computational modeling.

Intervention levels

According to the Interventions Wheel [23], the interventions could target three different levels, individual, community, and system. We suggest implementing our interventions in three ways that focuses on two levels of the Interventions Wheel: community and individual. First, community-based intervention is on the community level in the Intervention Wheel. Second, the school- and family-based interventions are on the individual level in the Intervention Wheel. Below, we provide examples of how to apply these sublevels for our purposes of mitigating social media challenge adoption.

Community-based intervention

Posters and handouts containing storyboards telling a similar story as the social media challenge yet highlighting the unwanted consequences and peer disapproval could be utilized. These posters could be advertised in public places such as parks and malls by county task forces. These methods have the potential of increasing the educational level about a certain behavior, but may not be as effective as other more interactive methods [27]. The content must be carefully created to avoid any contagion risk of media depiction of self-harming behavior [27].

School-based

Small discussion groups, videos, and peer-to-peer role-playing activities can be provided as a session along with existing educational programs such as Mothers Against Drunk Driving (MAAD) [28].

Family-based

An alternative and potentially more convenient way to implement the interventions is by sending the video to the parents to be played directly by them for their children.

Another way of eliminating participation in social media challenges is by detecting videos about those challenges and removing them from social media platforms to further reduce the spread of the challenge [16].

Computational modeling

Computational simulation is a feasible, flexible, and collaborative tool for modeling complex systems. One of its most widely used classes is agent-based modeling (ABM), which has been used extensively to investigate human behavior and complex phenomena as well as to explore complex problems in several domains, including the social sciences [29, 30] and healthcare [31, 32]. For example, Epstein [33] developed an ABM for investigating civil violence and rebellion. More relevant to the research reported here, it has been used to explore teenager behavior, specifically focusing on investigating the effects of various interventions on younger adults. For example, Yonas et al. [34] developed a conceptual ABM representing abstract community crime perpetrated by adolescents in addition to exploring the effects of several community-wide interventions on the contagion of these crimes [16].

Agent-based modeling provides deeper insight and more realistic outcomes about the effect of interventions, because it can be modeled to account for the possible consequences of an intervention. Because ABM offers a unique and cost-effective way to construct, assess, and implement a variety of behavioral interventions in a simulated dynamic environment, it has the potential to enhance understanding of the contagion effect of social media challenges and evaluate the potential efficacy of an intervention [35].

Due to the nature of the recommended interventions, we expect two opposing effects to emerge. First, when some people hear about a challenge, even through an educational program, they might still consider participation. In fact, they might have never heard about the challenge before, but the educational program may trigger them to consider researching it and subsequently participating. This response is referred to as the contagion effect [36]. On the other hand, these educational programs might have a larger positive impact than expected due to the word-of-mouth effect. That is, individuals who were successfully impacted by the education program and chose not to participate in harmful social media challenges are likely to discuss what they learned with others, in turn discouraging them from participating as well.

As a result, the goal of this study was to develop a conceptual ABM that represents the contagion dynamics of these social media challenges. This model includes the essential features of adolescents who participate in these challenges informed by the results from the previous study [12]. In addition, this ABM will lay the foundation for future models to further address this phenomenon.

Methods

The goal of our model is to build an individual-level ABM able to temporarily present the dynamics of the spread of a social media challenge in a small community. Additionally, through this ABM, we seek to test the effect of potential interventions and combinations of interventions at several levels including, family, school, and community. This section includes (1) the model and how the agents interact with one another; (2) the integration of the data obtained from the survey [12] into the model and how they affect the agent's interaction.

Model overview and design

The ABM was developed using Anylogic 8.5.1 University Researcher Edition [37] simulation software, a modeling tool that supports agent-based, discrete event, and systems dynamics methodologies using Java scripts. The simulated community included five agent types: child, father, mother, house, and school. A two-dimensional 100,000 × 100,000 toroidal grid was utilized to represent the community with one or more agents occupying any specific location. This was done to mimic a traditional neighborhood where each house could have neighbors around them from many directions. Each school and house in the environment were randomly located in the two-dimensional toroidal grid. Then, each child, father, and mother were randomly assigned to and located at one house and together formed a family. Children were connected to the closest school. Next, each child established a close relationship with the nearest child (a "best friend") along with a minimum of 5 and a maximum of 25 other friends from the same school.

The simulation time was set in days, and on each day, the agents updated their individual characteristics. Below are descriptions of agent updates based on agent type.

Child

The formation of the child's behavior is motivated by the integrated behavioral model (IBM) [38], which has one key element determining the child's decision to participate or not in a challenge: intention to participate. The primary key contributors to the intention to participate include the attitudinal components "attitude," the perceived norm components "norm," the personal agency components "personal agency," and other factors combined as "individual differences." As the model moves forward in time, these components of the children evolve based on their interactions with other agents (including their parents) in their social network and feedback from the environment, described later in this section. The decision criteria are based on the child's (agent's) intention to participate, which must be above a certain threshold before she/he participates in the social media challenge. Figure 1 presents the daily time-step of the child agents within the ABM. Each child begins forming beliefs about the challenge once one child in their social network has participated in it. Children are assigned individual initial values of experiential and instrumental attitudes, injunctive and descriptive norms, perceived control, and self-efficacy. Thus, children become more or less inclined to participate in the challenge depending on other children's experiences with the challenge and their interactions with their school and parents.

Parents (father and mother)

Each child is assigned to parents who could intervene and ask the child not to participate in the challenge if they know about it. A portion of the parent population will not know about the challenge and, thus, will not intervene. For those who intervene, if their child has not already participated in the challenge, then the probability of him/her participating will decrease. The value of the reduction in the probability of participation was determined using a trial-and-error method to establish a model that represents the collected data. We started with a %5 reduction in the probability of participation and compared the percentage of children who participated in the collected data using a *t* test. We increased or decreased the reduction in the probability of participation based on whether the mean percentage of children who participated in the simulated environment was lower or greater than those who participated in the collected data. We iterated this process until the difference was not statistically significant.

School

Similar to the parent population, schools' officials could also intervene and ask their students (children) not to participate in the challenge if they know about it. Similar to the parent population, a portion of the schools will not know about the challenge and, therefore, will not be able to intervene. For the schools that intervened, the probability of the students participating in the challenge who have not participated yet will decrease. The value of the reduction in the probability of participation was determined using a trial-and-error method to establish a model that represents the collected data. We started by running the model 63 times at 50% probability of participation and compared the percentage of children who participated in the collected data use a *t* test. We increased or decreased the probability of participation based on

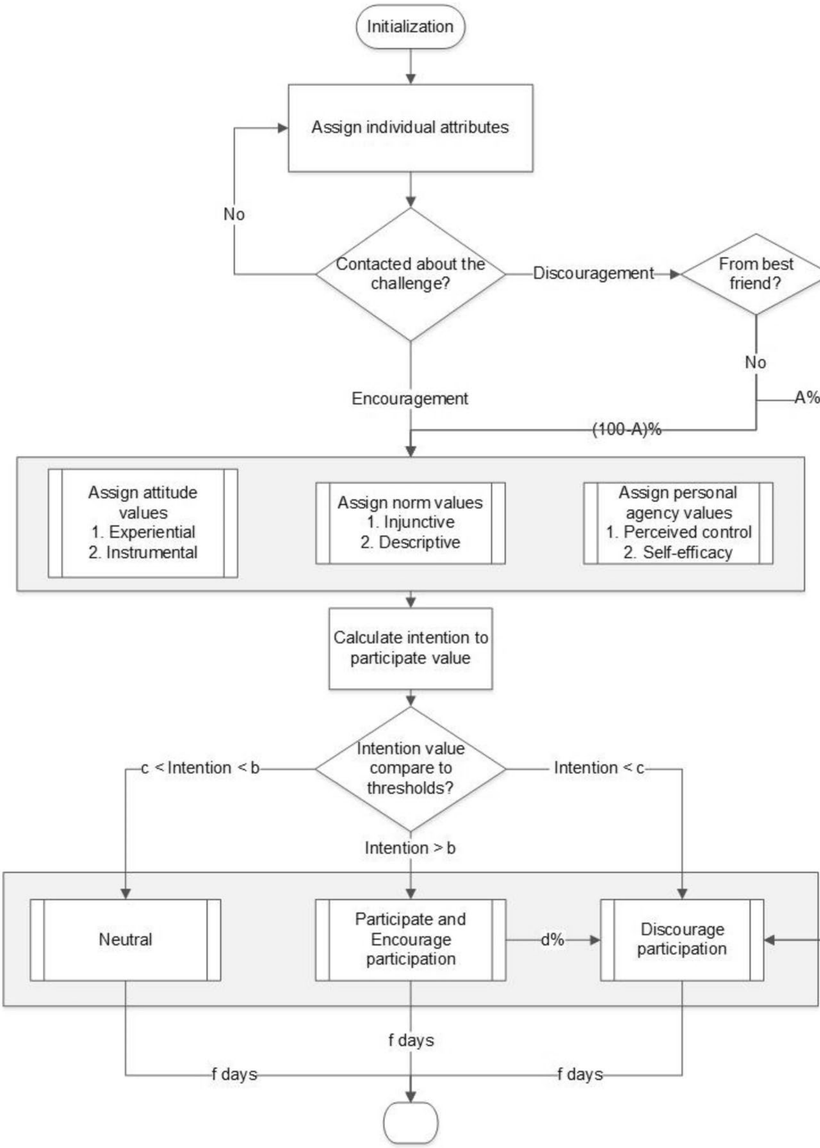

Fig. 1 The daily time-step of the child agents in the ABM

whether the mean percentage of children who participated in the simulated environment was lower or greater than those who participated in the collected data. We iterated this process until the difference was not statistically significant.

Home

The homes in the model had no effect on the decision to participate. They were included in the model for the purpose of forming families at one location. The animations of all agent types are presented in Fig. 2.

Data

Data collection and description are reported in a previous publication [12]. To develop this model, we only included data from those who participated in the

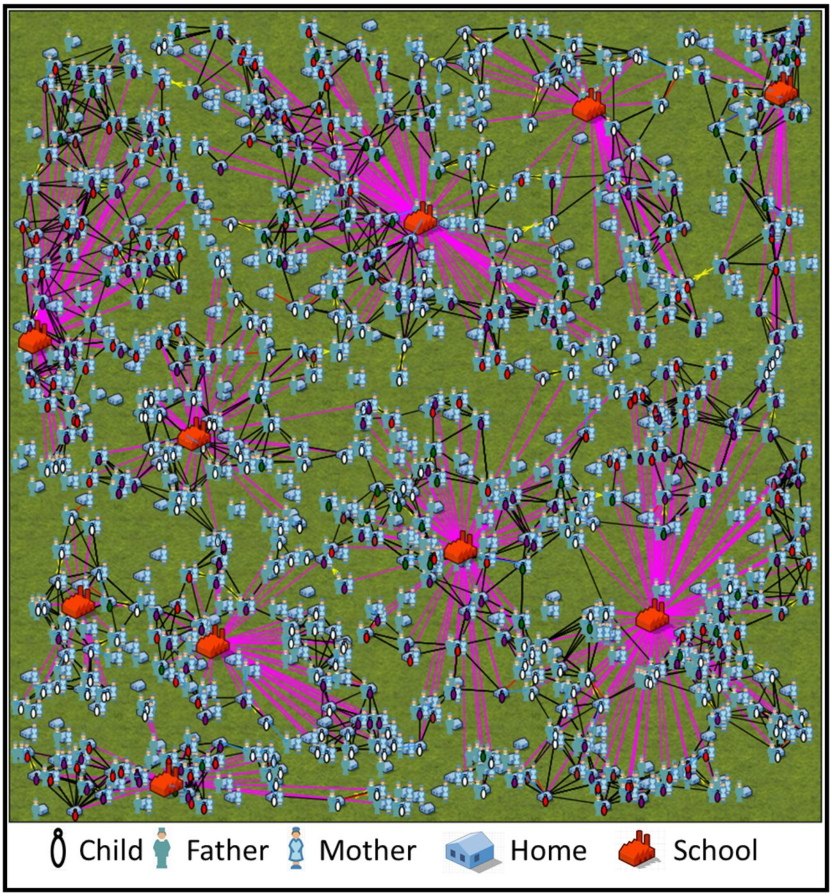

Fig. 2 A close-up view of the agents in the ABM. *Note*: The color of a child indicates his/her status: children who have not heard about the challenge are white; those who considered participating are purple; those who participate are red; and those who discourage participation are green. The color of the line indicates the type of relationship between the agents: friends are connected by black lines; best friends are connected by yellow lines; and children are connected to their fathers by blue lines, to their mothers by red lines, and schools by pink lines

Cinnamon Challenge. All the participants were 18 years old or younger at the time of participation in the challenge, although some were older than 18 years old at the time they took the survey. The distributions of the attitudinal components “experiential and instrumental attitude,” the perceived norm components “injunctive and descriptive norm,” the personal agency components “perceived control and self-efficacy,” and other factors “individual differences” were based on the data provided in our prior study [12], and were tested for normality using their skewness and kurtosis values [39]. All data were normally distributed. Initial values for these factors were randomly assigned to each child in the model based on the most fitting mean and standard deviation with a minimum value of 1 and a maximum of 7.

To determine the values for intention to participate for each child, we conducted a linear regression using the data collected from the survey in [12] with the intention to participate in the Cinnamon Challenge measured on a 7-point Likert scale being the dependent variable and the factors introduced previously as the independent variables. The linear regression established that instrumental attitude, injunctive norm, descriptive norm, and impulsivity statistically significantly predicted intention to participate in the Cinnamon Challenge, $F(6, 153) = 153.318$, $p < 0.001$. The regression equation with the significant predictors only is provided below. In the model, we included all the factors, even the non-significant predictors to calculate the intention to participate. However, only the significant ones had a high impact on the calculated intervention to participate value as their slopes were significantly higher than the rest of the predictors.

$$\begin{aligned}\text{Intention} = &- 0.70 + 0.36 \times \text{(instrumental attitude)} \\ &+ 0.26 \times \text{(injunctive norm)} + 0.37 \times \text{(descriptive norm)} \\ &+ 0.14 \times \text{(impulsivity)}.\end{aligned} \tag{1}$$

Calibration and validation

To validate the model, we used a multistage validation technique [40]. After developing a model based on theory, observations, and knowledge of the behavior, its assumptions must be validated by empirically testing and comparing them to data collected from a real system. To calibrate the model and capture trends in our model similar to the survey reported in [12], we developed parameterized rules and assumptions to change the participation probabilities of the children:

- R1: The child will participate if his/her intention to participate > participation threshold.
- R2: The child will discourage others from participating if his/her intention to participate < the discouraging threshold or if they are discouraged to participate by their best friend.

- R3: A portion of the children who participated will discourage participation (those who had a negative experience with the challenge), while the rest will encourage participation (those who had a positive experience).
- R4: The child will stop discouraging or encouraging participation after a certain number of days.

The values of the parameters for these rules (participation threshold, discouraging threshold, portion of children having a negative experience, and number of days to stop encouraging/discouraging other children) were calibrated by trial-and-error, so that the resulting behavior of the children in the model matches the sample we collected in [12]. We run the model 63 times, compare the results to the observed data, and make necessary adjustment to the model parameters to try to get closer to the values in the collected data. We repeated this process until there was no statistically significant difference between the number of children who participated from the model and the collected data. That is, the number of people who participated in the survey falls within the 99% confidence interval of the number of people who participated in the model after N runs. A one-sample t test showed that the number of children who participated over 63 replications of the simulation was not statistically significantly different ($M=144.41$, SD = 27.28) than the number of participants from the survey, which was 153 participants, $t(63)=-1.625$, $p=0.11$. In addition, an independent-samples t test was conducted to determine if there were differences in intention to participate between the survey data and the simulation outputs. The intention to participate was not statistically significant [$M=0.10$, 95% CI (−0. 14, 0.36), $t(926)=0.82$, $p=0.97$] between the survey data ($M=3.06$, SD = 1.99) and the ABM outputs ($M=2.95$, SD = 1.94).

Visualization and verification

A verification method widely used in simulation models is animation [41]. This Anylogic software provides a dynamic visualization of the model as it is running, providing validation that the computational implementation is in agreement with the conceptual model. A close-up view of the agents' interactions in the model can be seen in Fig. 2. In addition, we used the structured walkthrough approach [42] where the developer formally presented the model to two other researchers to determine its correctness. Then, the model was analyzed statically by the developer and each of the two researchers individually to determine whether it is correct based on the structured walkthroughs, the correctness proofs, and the examination of its structural properties.

Modeling of interventions

The interventions target three different levels, community school, and family. To determine the impact of each level on the child's attitude, perceived norm, and personal agency values, we conducted a within-subject ANOVA with the motivation to comply factor as the dependent variable and the relationship with the individual as the independent variable. We used within-subject ANOVA as the motivation to comply with the different norm categories that were measured for all the participants. The motivation to comply was significantly different for different relationship types, $F(1.94, 913.54) = 90.17$, $p < 0.001$, partial $\eta^2 = 0.16$, with motivation to comply decreasing from family ($M = 5.01$, SD = 1.79) to most people ($M = 4.49$, SD = 1.90) and then to friends ($M = 3.82$, SD = 1.88). Post hoc analysis with a Bonferroni adjustment revealed that motivation to comply significantly decreased from family to most people [$M = 0.52$, 95% CI (0.31, 0.72), $p < 0.001$], and from family to friends [$M = 1.19$, 95% CI (0.96, 1.42), $p < 0.001$], and from most people to friends [$M = 0.67$, 95% CI (0.46, 0.87), $p < 0.001$]. Based on these findings, we assumed that the family-based interventions will have the highest reduction in children's intentions to participate. As suggested by the literature, we assumed that school-based interventions have more of an impact on intention to participate than the community-based ones [23, 43]. We assumed the specific values of the reduction in intention to participate at each intervention level. We also assumed that it takes three days after the simulation starts for the interventions to be active. Below is a description of how we modeled these interventions in the ABM.

Family level

This intervention targets only parents who did not know about the existence of the challenge and did not intervene at the start of the simulation. We focused on how changes in a specified percentage of parents will affect the overall participation pattern. We assumed that if the parents received an intervention, then they will become active and ask their child not to participate. We assumed that if their child has not participated in the challenge yet, she/he will get attitudinal, perceived norm, and personal agency component values that are on average 0.5 less than those who did not receive the intervention.

School level

Similar to the family-level interventions, we assumed that if a school received an intervention, then those students who go to that school will get attitudinal, perceived norm, and personal agency component values that are on average 0.2 less than those who did not receive the intervention.

Community level

A certain percentage of the child population from the entire population was randomly selected to receive the intervention. If the child had not yet participated in the challenge, she/he will receive attitudinal, perceived norm, and personal agency component values that are on average 0.1 less than those who did not receive the intervention.

The ABM we built is a stochastic simulation resulting in different output and observations for each run of the model. Thus, several replications were run to gather statistics and evaluate the various interventions recommended. We used the confidence interval method recommended by Law [37] seen in the equation below to calculate the required number of replications per run, where h is the half width of the confidence interval and s is the standard deviation

$$\text{Number of replications} = Z^2_{1-\alpha/2}\frac{s^2}{h^2}. \quad (2)$$

To determine the initial standard deviation, we conducted a pilot run of the model for 20 replicates with a population size of 30,000 and calculated the standard deviation for the number of children who participated in each replication and found it to be 103. We specified precision in terms of the half width of the confidence interval to be 25 and an alpha level (α) of 0.05 ($Z = 1.96$). Using the equation above, we found the number of replications required to meet our precision criteria to be 63 replications. We ran the model using Anylogic on a 64-core machine, running a 64-bit Windows 2010 Enterprise with 32 GB of RAM and 3 GHz Intel Core i7-9700 processors. We determined the time required for the model to reach the steady state (no variation in the results) based on the number of children who participated and found it to be ~ 100 days. Therefore, we determined the stop time for each run to be 200 days as a conservative choice. The total running time for each run with 63 replications and a population of 30,000 was approximately 70 min. No warm-up period was needed.

Results

In all three intervention levels, we defined the intensity of the intervention as the percentage of each agent type (family, school, or community) that received the intervention. We ran the model four times (each run with 63 replicates) for each intervention type, each with a different intensity and compared them to the base model without interventions. The four intensity levels for the interventions were low = 25%, medium = 50%, high = 75%, and all = 100%. For example, in the low-intensity family-level intervention, 25% of the families in the model received the intervention. Using fixed intensity levels, we were able to explore the different effects of the types of recommended interventions that require almost the same resources but which deployed them in a different way. In addition, we were able to address the contagion effects resulting from those interventions. That is, at which level of intervention, the “word of mouth” effect will help in spreading these interventions further into the community.

The impact of individual interventions

We used a one-way ANOVA to compare the means of the number of children who participated after applying each intervention over 63 replications at each level. There were no outliers, as assessed by boxplot; data were normally distributed for

Table 1 The means and standard deviations of the number of children who participated after applying each intervention

Intervention level	Base model	Family level	School level	Community level
Low	3149 (137)	**2669 (114)****	**2919 (177)***	**3020 (119)**
Medium	3136 (134)	***2210 (102)*****	***2719 (189)****	***2871 (144)***
High	3136 (134)	***1783 (88)*****	***2521 (168)****	***2769 (143)***
All	3136 (134)	***1455 (101)*****	***2293 (100)****	***2648 (125)***

In each row, the means and standard deviations for the three types of interventions are presented in boldface if they show a statistically significant ($p < 0.001$) decrease compared to the base model. They are marked with a single asterisk if they show a statistically significant decrease when compared to one alternative intervention and a second asterisk if they show a significant decrease when compared to two alternative interventions. They are italicized if they show a statistically significant decrease when compared to a lower level of the same intervention

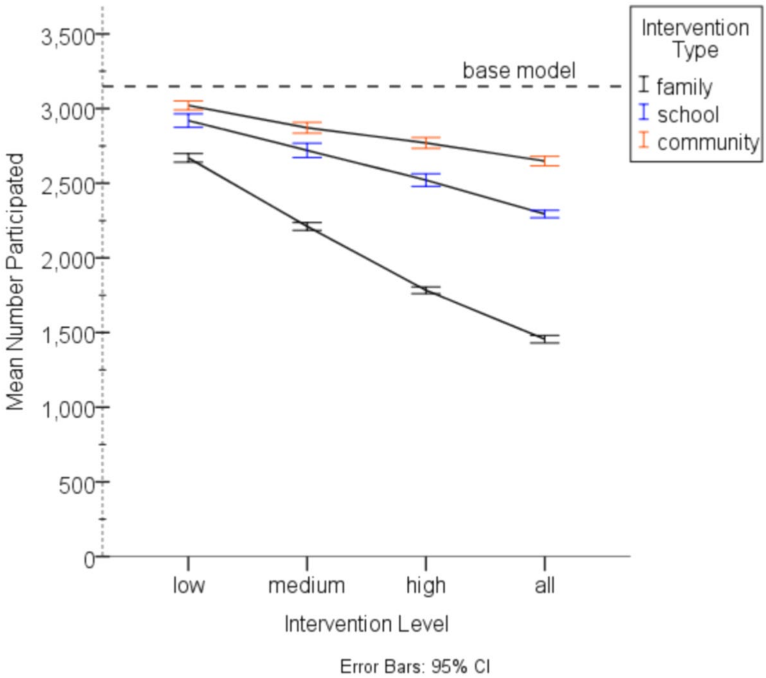

Fig. 3 Number of children who participated in the challenge after applying each intervention

each group, as assessed by the Shapiro – Wilk test ($p > 0.05$); and variances were homogeneous, as assessed by Levene's test of homogeneity of variances [44]. Data are presented as mean ± standard deviation in Table 1 and plotted in Fig. 3.

Sensitivity analysis

We included multiple parameters in our model for which the data are currently unavailable. Therefore, it is necessary to explore the sensitivity of our results to these parameters. One important parameter is the percentage of children initially participating in the social media challenge as it affects the probability of a child hearing about the challenge and considering his/her own participation, thus affecting the final number of children who participated. Furthermore, this value can also affect the probability of a child initially not liking the challenge and, therefore, discouraging other children from participation, resulting in a decrease in the effectiveness of the intervention. To ascertain the sensitivity of the findings to the percentage of children who initially participated, variations of the models were created with 0.5%, 1%, 2%, 3%, and 4% of the children initially participating. These percentages were selected by comparing the percentage of participation to those obtained from the survey. We ran these variations for 63 replications only on three levels (low, medium, and high) of the family, school, and communitybased interventions since all the other levels will follow a similar trend. The results are presented in Figs. 4, 5, 6, 7, 8 and 9. For all models, family based interventions reduce the number of children who participated more than the school- and community-based interventions. Overall, the performance of all types of interventions was consistent with the tested variations in the percentage of children who initially participated.

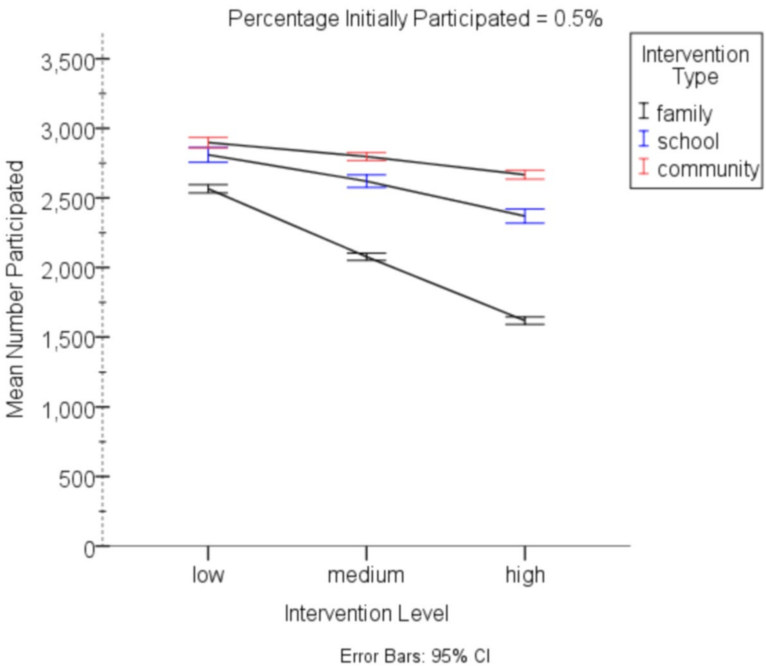

Fig. 4 Effects of interventions when the percentage of children who initially participated is 0.5%

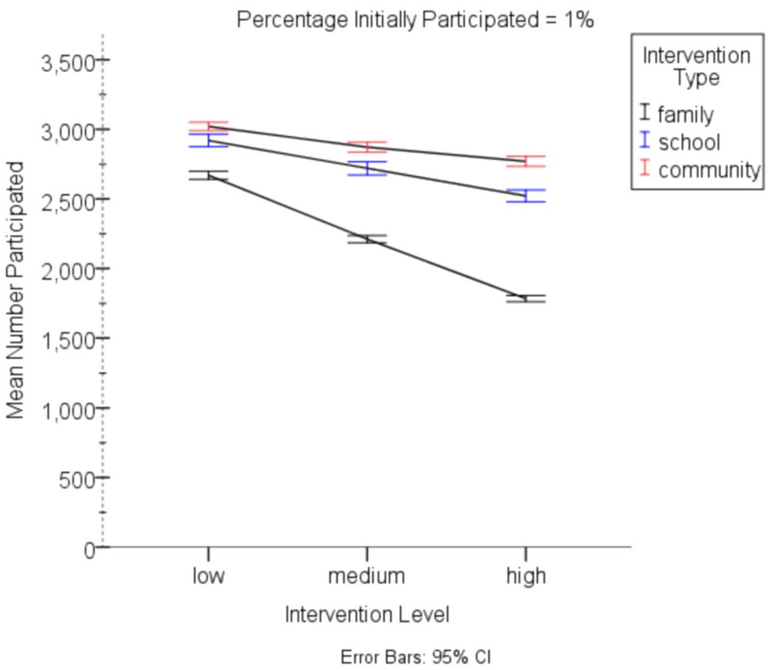

Fig. 5 Effects of interventions when the percentage of children who initially participated is 1%

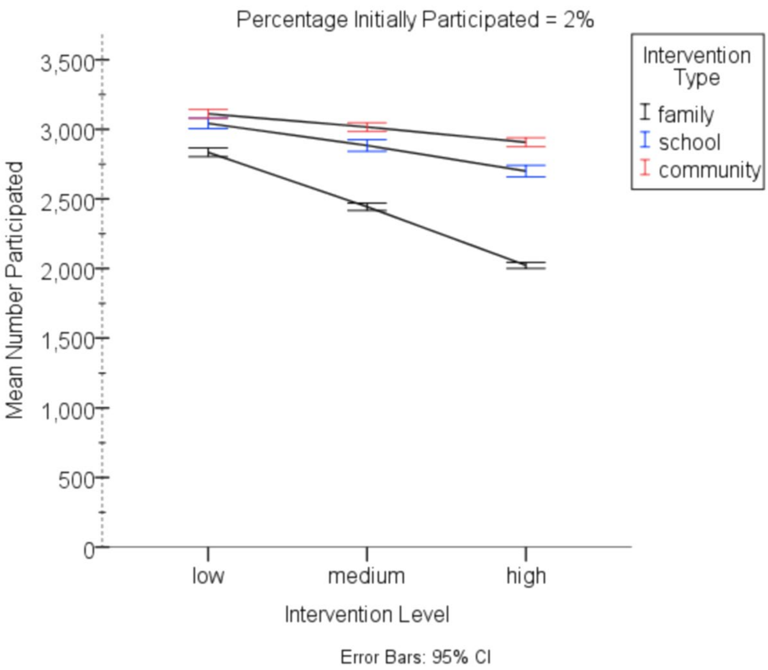

Fig. 6 Effects of interventions when the percentage of children who initially participated is 2%

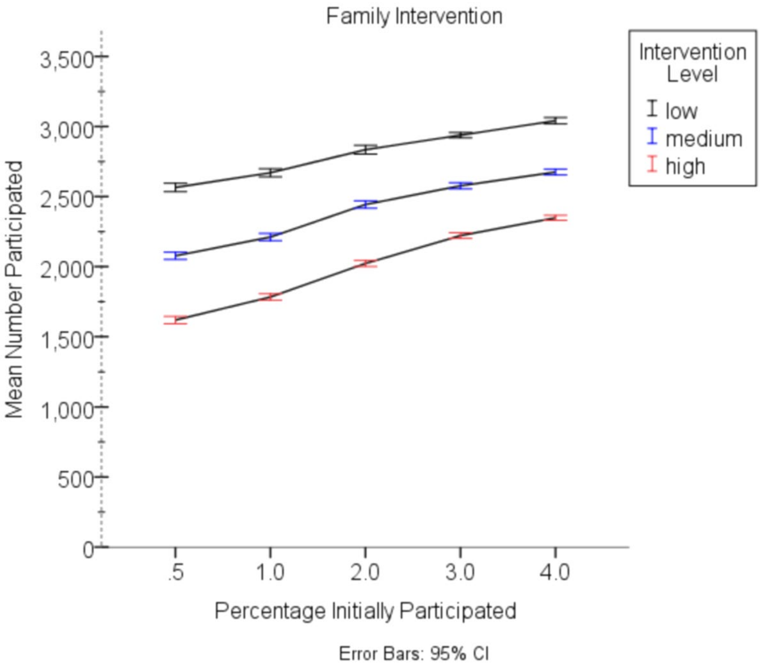

Fig. 7 The effect of percentage of children who initially participated in a family-based intervention

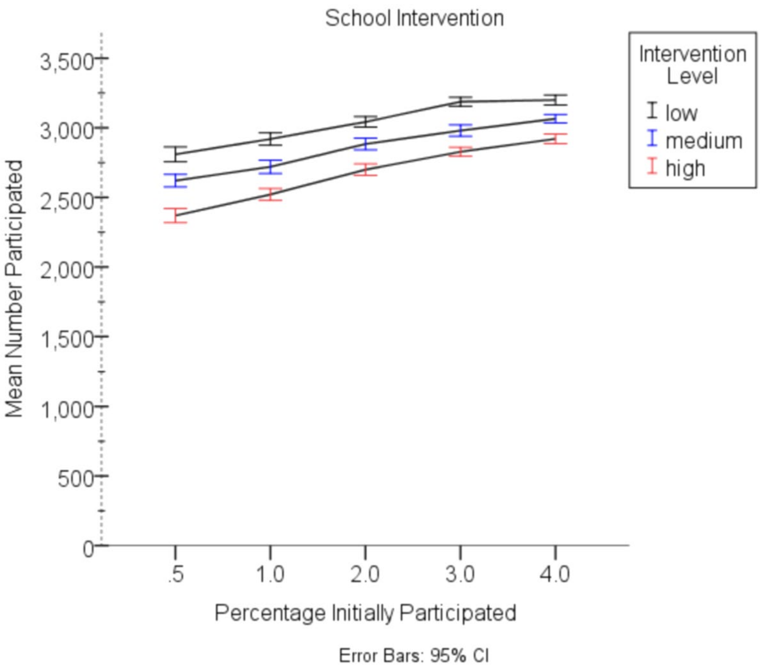

Fig. 8 The effect of percentage of children who initially participated in a school-based intervention

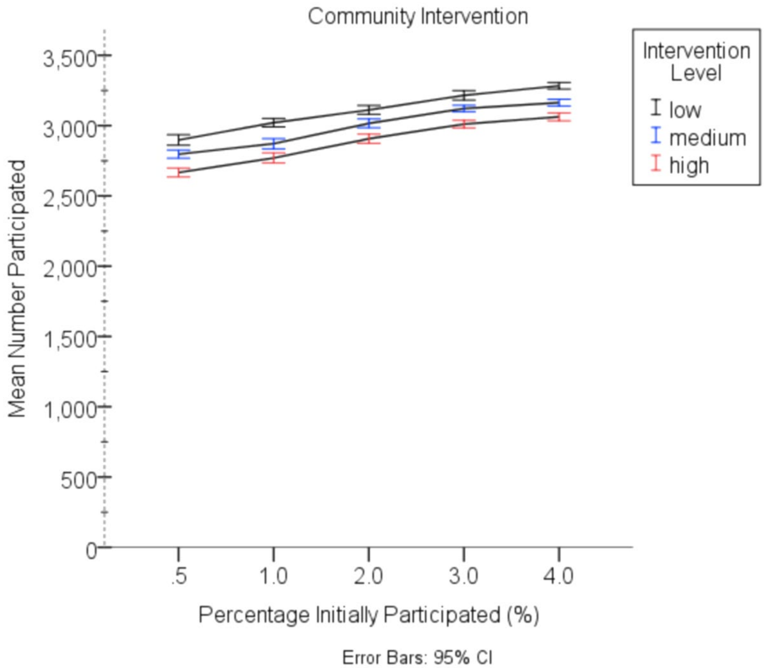

Fig. 9 The effect of percentage of children who initially participated in a community-based intervention

The impact of combinations of interventions

Because these results indicated that the school-based and family-based interventions showed significantly more reduction in the number of children who participated in a challenge than the community-based intervention, we decided to explore the effect of a combination of family- and school-based interventions on the number of participants. To do so, we ran the model 63 times for each

Table 2 The means and standard deviations for the number of children who participated after applying combinations of interventions

Family/school	Low	Medium	High	All
Low	2462 (144)	**2321** (**153**)*	**2096** (**171**)*	**1917** (**109**)*
Medium	**2068** (**153**)*	**1917** (**153**)*	**1752** (**159**)*	**1569** (**91**)*
High	**1651** (**113**)*	**1535** (**139**)*	**1404** (**114**)*	**1311** (**84**)*
All	**1353** (**116**)*	**1250** (**131**)*	**1129** (**122**)*	**1059** (**101**)*

In each row, the means and standard deviations for the combination of two types of interventions are presented in boldface if they show a statistically significant ($p < 0.001$) decrease when compared to the model that has one-level lower school intervention (the cell to the left of it). They are marked with a single asterisk if they show a statistically significant decrease when compared to the model with one-level lower family intervention (the cell above it)

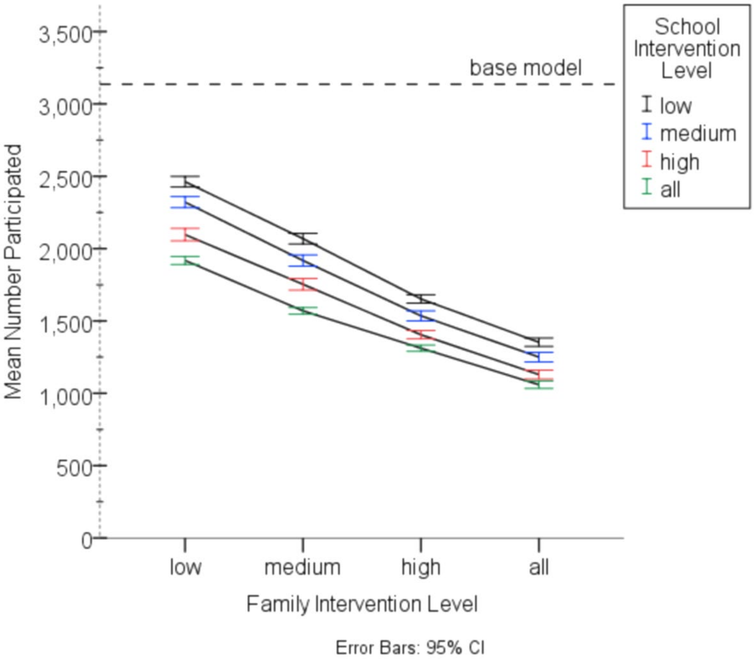

Fig. 10 Number of children who participated in the challenge after applying combinations of interventions

combination of the different levels of family-based and school-based interventions, resulting in 16 combinations.

We used a one-way ANOVA to compare the means of the number of children who participated after applying each combination of interventions over 63 replications at each level. Data are presented as mean ± standard deviation in Table 2 and plotted in Fig. 10.

Educational interventions versus social media-based policy interventions

Another way of eliminating participation in social media challenges is by detecting videos about those challenges and removing them from social media platforms to reduce the further spread of the challenge. We modeled these interventions in our ABM by reducing the percentage of children who encourage other children to participate. We assumed that this percentage is related to the accuracy of the detection algorithm. For example, if the algorithm exhibits an 80% accuracy, then only 20% of the children who participated and liked the challenge will encourage participation. We tested this policy-based intervention at two levels, 80% and 90% accuracy. Since the most recent algorithms in video detection exhibit accuracies around 90% [45], we selected it as a representative value. However frequently, these accuracy values are inflated because of bias in testing the algorithm using data similar to the training data and/or by consistently misclassifying adversarial examples [46]. In other words, the performances of these algorithms when applied in the real world are much lower than what is reported in the research [47]. Therefore, we also selected 80% accuracy as a more conservative choice. The results are presented in Fig. 10. The only combinations of interventions that significantly reduced ($p < 0.001$) the number of children who participated in the challenge beyond what an algorithm with 80% detection accuracy reduced were (1) the medium level of family-based intervention and the high level of school-based intervention [184, 95% CI (102, 267)] and (2) the medium level of family-based intervention and all schools receive intervention [367, 95% CI (284, 450)].

Discussion

Computational simulation methods such as ABM can help explore the effects of multiple interventions at varying levels to reduce participation in harmful behaviors. ABM in this research provided an effective means for the conceptual dynamic simulation of participation in harmful social media challenges and the potential impact of various multilevel interventions aimed at decreasing their virality. Although this is an abstract model, it reveals interesting trade-offs between the different interventions. The family-based interventions were the most effective, followed by the school-based interventions, then the community-based ones. When compared to video detection and deletion algorithms, the combination of family and school-based interventions were better at reducing participation only at high levels of interventions.

The impact of individual interventions

In general, the family-based interventions reduced participation more than all the other interventions. This finding is primarily because parents had the largest impact on children's decisions in the model. When the intervention is given to children by their parents, their intention to participate will be reduced more than if the intervention were delivered by anyone else, thus reducing the number of people who participated. Moreover, those who received the intervention from their parents are more likely to exemplify a very low intention to participate, leading them to discourage other children from participating, reducing participation further. This finding is consistent with most of the recent intervention literature that suggests involving families in interventions, especially those for children and adolescents [48].

In addition to family-based interventions, we found that school-based interventions reduced participation only slightly more than community-based interventions when the intervention level was low. However, at higher levels of interventions, school-based interventions were much better at reducing participation than community-based interventions. This finding was primarily due to the way interventions are deployed. When the school-based intervention is low, meaning only a few schools receive the intervention, its main effect and its word-of-mouth effect are localized to areas surrounding the school. However, the spread of community-based interventions among the children is wider than the school-based intervention when the level is low. The low community-based intervention will reduce the intention to participate in these challenges for people at different locations and its word-of-mouth effect will help spread it farther to other locations. On the other hand, when the intervention level is high, most people will receive the direct effect of the intervention, reducing the word-of-mouth effect as it will most likely get to those who have already received an intervention and developed a low intention to participate. This finding is consistent with earlier literature on the contagion effect of community-wide interventions [34, 49]. For example, Yonas et al. [34] tested the contagion effect of community-wide vs. spatial-focused crime reduction interventions, finding that community-wide interventions are more contagious than spatially focused ones at low levels of interventions.

The impact of combinations of interventions

We conducted several combinations of educational interventions at the family and school levels to explore the added effect of additional interventions on reducing participation. We found that adding another intervention type significantly reduced participation. In addition, increasing family-based interventions, one level (within the same type of intervention) was more effective than increasing one level of school-based interventions. However, increasing the school-based intervention by two levels had the same impact or better than adding one level of school-based intervention. Overall, the effect of increasing the level of intervention is reduced as the intervention level increases. For example, if a high percentage of the population receives a family-based intervention, then providing a higher level of a school-based intervention will not be as valuable as when the family-based intervention is low. This result is primarily due to the increased probability of providing the intervention to children who already received an intervention or have a low intention to participate. This common trend in the literature led to the development of focused interventions, which target only those with a high probability of engaging in a behavior when possible [34]. Therefore, as a future direction, our interventions could focus only on those with higher impulsivity instead of everyone in the community as a more efficient alternative.

Our model could help decision-makers assess the availability of resources and decide on which intervention type, combination, and level are more appropriate than others. For example, family-based interventions may be more cost-effective than other types as the delivery method relies on parents, thus requiring fewer resources. If this is the case, then it becomes more beneficial to choose family-based interventions over others. The school-based interventions could lead to consistent and sustained reductions in participation in harmful social media challenges if resources are available for high-level interventions and the intervention is provided to many schools. However, when resources are limited, community-based interventions may have a better impact on reducing participation. The added value of a combination of interventions can also be assessed. For instance, if the resources required to increase a family-based intervention is the same as providing the intervention through a few more schools, then the latter option might be more efficient as it is likely to have a greater impact on participation.

Educational interventions versus social media-based policy interventions

We compared the educational-based interventions at the family and school levels with the policy-based interventions by detecting and removing all videos from social media platforms at two levels of accuracy: 80% and 90%. In general, the policy-based intervention with 80% accuracy reduced participation more than any single intervention by itself. The combination of interventions reduced participation more than the policy-based intervention with 80% accuracy only when there were at least medium-level interventions of both family and school-based or low school-based with high or all family-based interventions. When the detection accuracy was 90%, only combinations of interventions with high levels or more of both family and school-based interventions were more effective at reducing participation. These results highlight interesting trade-offs on which type of intervention will reduce participation more than others to help decision-makers make effective choices about which option to use. While accuracy detection algorithms might seem to be an effective and efficient option in reducing participation, one should consider that most of these algorithms rarely reach 90% accuracy [46] when implemented, thus possibly making them less effective than educational methods. In addition, over the long term, the educational-based intervention methods are more efficient as they are more likely to target any type of social media challenge and reduce any future participation in those challenges, while on the other hand, algorithms have to be developed each time a new challenge appears. In other words, educational interventions target the root of the problem, while detection algorithms act as a retrospective solution.

ABMs have been used in the literature to explore the dynamics of multiple unwanted behaviors such as crime reporting interventions for criminal behaviors focused on the community level [34]. Although this has expanded the usability of ABMs, there is a lack of research focused on comparing multiple levels, combinations, and types of interventions. Our model is novel in the application field it used, social media challenges, and its focus on multiple levels of educational interventions, combinations of interventions, and comparison with policy-based intervention. Although interpretation is limited due to the unavailability of certain data, we will evolve the model in the future to be more robust and empirically informed. In addition, the model provides the ability to test the model results in minutes, making it a practical alternative to the much more time-consuming, expensive, and resource-consuming traditional methods of testing alternative interventions.

Limitations and future work

As with any research study or simulation model, there are a number of limitations we need to mention. This ABM is an early conceptual model that represents abstract dynamic interactions related to social media challenges focused on interactions between children, their peers, schools, and parents. Although this model is partially empirically driven, as noted earlier, there are some limitations related to the unavailability of certain data such as demographics, social networks, and the actual effect and cost of varying interventions on the intention to participate. Future interactions of this model will, therefore, include a baseline for decision-making on targeting interventions for children focused not only on reducing participation in social media challenges but also on reducing any other unwanted behaviors. We will also examine the dynamics of the spread of challenges to determine the appropriate timing of introducing an intervention. Given the current popularity of the social media, more detailed empirically grounded models are needed to model specific levels of participation in the challenges posted on these sites.

Additionally, future directions of this research will seek to investigate the contagion effect of not only social media challenges but also suicide and other means of self-harm. We will test the effect of other social-based policy interventions such as deforming the "social bubbles" of adolescents with suicidal thoughts on reducing the contagion risks of self-harm and suicide. Given the strong correlation between collective and individual diversity, people find themselves inside "social bubbles" when they use social media [50], This is either due to people's tendency to selectively expose themselves to the opinion of other like-minded individuals [51, 52] or the effect of the personalized filters (algorithms) implemented by social media platforms, such as Facebook, to expose the user to posts only by like-minded individuals [53]. We hope that deforming the social bubbles of those with suicidal thoughts and exposing them to others who are less like-minded will reduce the contagion risks of self-harm and suicide at a societal level. We also believe that doing so will enhance compliance with the Safe and Effective Messaging Guidelines as this strategy will reduce the exposure of vulnerable individuals to self-harm content and provide them with more help-seeking opportunities [54].

Conclusion

In this work, we showed how ABM might be used to investigate the effect of educational intervention programs to reduce the participation in social media challenges at multiple levels, specifically family, school, and community. In addition, we showed how the effect of these educational-based interventions can be compared to social media-based policy interventions. This model accounted for the word-of-mouth effect of these interventions, finding that it could either decrease participation in social media challenge further than expected or unintentionally cause others to participate. We suggest that educational interventions at a combined family and school level are the most effective types of intervention in the long run. Future plans include testing the effect of educational interventions on the intention to participate in a quasi-experimental study and incorporating these values into a more sophisticated model that can help inform the design and selection of future intervention programs.

Acknowledgements This material is based on work supported by the National Science Foundation under Grant No. 1832904.

Data availability statement The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest All authors declare that they have no conflict of interest.

References

1. FastStats—Suicide and Self-Inflicted Injury. (n.d.). https://www.cdc.gov/nchs/fastats/suicide.htm (accessed July 20, 2021).
2. Twenge, J. M., Joiner, T. E., Rogers, M. L., & Martin, G. N. (2018). Increases in depressive symptoms, suicide-related outcomes, and suicide rates among US adolescents after 2010 and links to increased new media screen time. *Clinical Psychological Science., 6*, 3–17. https://doi.org/10.1177/2167702617723376
3. NIMH » Suicide. (n.d.). https://www.nimh.nih.gov/health/statistics/suicide (accessed July 20, 2021).
4. Mitchell, K. J., Wells, M., Priebe, G., & Ybarra, M. L. (2014). Exposure to websites that encourage self-harm and suicide: Prevalence rates and association with actual thoughts of self-harm and thoughts of suicide in the United States. *Journal of Adolescence, 37*, 1335–1344. https://doi.org/10.1016/j.adolescence.2014.09.011
5. Zhu, L., Westers, N. J., Horton, S. E., King, J. D., Diederich, A., Stewart, S. M., & Kennard, B. D. (2016). Frequency of exposure to and engagement in nonsuicidal self-injury among inpatient adolescents. *Archives of Suicide Research, 20*, 580–590. https://doi.org/10.1080/13811118.2016.1162240
6. Pater, J., Mynatt, E. (2017). Defining digital self-harm. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing—CSCW'17* (pp. 1501–1513). ACM Press. https://doi.org/10.1145/2998181.2998224
7. Rachmi, C. N., Agho, K. E., Li, M., & Baur, L. A. (2016). Stunting, underweight and overweight in children aged 2.0–4.9 years in Indonesia: Prevalence trends and associated risk factors. *PLoS ONE, 11*, e0154756. https://doi.org/10.1371/journal.pone.0154756
8. Luxton, D. D., June, J. D., & Fairall, J. M. (2012). Social media and suicide: A public health perspective. *American Journal of Public Health, 102*(Suppl 2), S195–S200. https://doi.org/10.2105/AJPH.2011.300608
9. Grant-Alfieri, A., Schaechter, J., & Lipshultz, S. E. (2013). Ingesting and aspirating dry cinnamon by children and adolescents: The "cinnamon challenge." *Pediatrics, 131*, 833–835. https://doi.org/10.1542/peds.2012-3418
10. 48-Hour Challenge: Is the Disappearing "Game" a Hoax? | Miami Herald. (n.d.). https://www.miamiherald.com/news/nation-world/national/article180181516.html (accessed July 20, 2021).
11. Social Media Contests: A Timeline of Viral Challenges. (n.d.). https://growingsocialmedia.com/social-media-contests-a-timeline-of-viral-challenges/ (accessed July 20, 2021).
12. Khasawneh, A., Madathil, K. C., Zinzow, H., Rosopa, P., Natarajan, G., Achuthan, K., & Narasimhan, M. (2021). Factors contributing to adolescents' and young adults' participation in web-based challenges: Survey study. *JMIR Pediatrics and Parenting, 4*, e24988. https://doi.org/10.2196/24988
13. Khasawneh, A., Madathil, K. C., Zinzow, H., Wisniewski, P., Ponathil, A., Rogers, H., Agnisarman, S., Roth, R., & Narasimhan, M. (2021). An investigation of the portrayal of social media challenges on youtube and twitter. *Transactions on Social Computational, 4*, 1–23. https://doi.org/10.1145/3444961

14. ALS Ice Bucket Challenge Commitments | The ALS Association. (n.d.). https://www.als.org/ice-bucket-challenge-spending (accessed July 20, 2021).
15. Khasawneh, A., Madathil, K. C., Dixon, E., Wisniewski, P., Zinzow, H., & Roth, R. (2019). An investigation on the portrayal of blue whale challenge on Youtube and twitter. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting., 63*, 887–888. https://doi.org/10.1177/1071181319631179
16. Khasawneh, A., Madathil, K. C., Dixon, E., Wiśniewski, P., Zinzow, H., & Roth, R. (2020). Examining the self-harm and suicide contagion effects of the blue whale challenge on YouTube and Twitter: Qualitative study. *JMIR Mental Health, 7*, e15973. https://doi.org/10.2196/15973
17. Roth, R., Abraham, J., Zinzow, H., Wisniewski, P., Khasawneh, A., & Madathil, K. C. (2020). Evaluating news media reports on the "blue whale challenge" for adherence to suicide prevention safe messaging guidelines. *Proceedings of the ACM on Human–Computer Interaction, 4*, 1–27. https://doi.org/10.1145/3392831
18. Tide Pod Challenge: Tide Tells People to Call Poison Control | Time. (n.d.). https://time.com/5108478/tide-pod-challenge-poison-control/ (accessed July 20, 2021).
19. P.M.C. Europe, Europe PMC. (n.d.).
20. Khasawneh, A., Narasimha, S., Madathil, K. C., Zinzow, H., & Rosopa, P. (2020). An investigation of the factors predicting participation in social media challenge. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting., 64*, 810–811. https://doi.org/10.1177/1071181320641188
21. Mukhra, R., Baryah, N., Krishan, K., & Kanchan, T. (2019). "blue whale challenge": A game or crime? *Science and Engineering Ethics, 25*, 285–291. https://doi.org/10.1007/s11948-017-0004-2
22. Agnisarman, S., Madathil, K. C., Khasawneh, A., Ponathil, A., Lopes, S., Piratla, K., & Gajjar, D. (2022). Sensemaking perspective on infrastructure risk-related mental model development of windstorm loss prevention engineers. *Journal of Building Engineering., 52*, 104314. https://doi.org/10.1016/j.jobe.2022.104314
23. Publications Database—Center for Public Health Practice—Minnesota Dept. of Health. (n.d.). https://www.health.state.mn.us/communities/practice/resources/publications/ (accessed July 20, 2021).
24. Ponathil, A. P., Khasawneh, A., Byrne, K., & Madathil, K. C. (2020). Factors affecting the choice of a dental care provider by older adults based on online consumer reviews. *IISE Transactions on Healthcare Systems Engineering*. https://doi.org/10.1080/24725579.2020.1854394
25. Friedman, J., Cosby, R., Boyko, S., Hatton-Bauer, G., Turnbull, Panel, T. (n.d.). *Effective teaching strategies and methods of delivery for patient education*.
26. Friedman, A. J., Cosby, R., Boyko, S., Hatton-Bauer, J., & Turnbull, G. (2011). Effective teaching strategies and methods of delivery for patient education: A systematic review and practice guideline recommendations. *Journal of Cancer Education, 26*, 12–21. https://doi.org/10.1007/s13187-010-0183-x
27. Keller, L. O., Strohschein, S., Lia-Hoagberg, B., & Schaffer, M. A. (2004). Population-based public health interventions: Practice-based and evidence-supported. Part I. *Public Health Nurse, 21*, 453–468. https://doi.org/10.1111/j.0737-1209.2004.21509.x
28. MADD | Mothers Against Drunk Driving. (n.d.). https://www.madd.org/ (accessed July 20, 2021).
29. Wang, S., Wang, J., Wang, X., Qiu, T., Yuan, Y., Ouyang, L., Guo, Y., & Wang, F.-Y. (2018). Blockchain-powered parallel healthcare systems based on the ACP approach. *IEEE Transactions on Computational Social Systems, 5*, 942–950. https://doi.org/10.1109/TCSS.2018.2865526
30. Ponathil, A., Khasawneh, A., Piratla, K., Hegde, S., Sharma, V., & Madathil, K. C. (2021). Does construction service provider's response matter? Understanding the influence of anecdotal information on online consumer decisions. *Developments in the Built Environment*. https://doi.org/10.1016/j.dibe.2021.100061
31. Lee, B. Y., Brown, S. T., Bailey, R. R., Zimmerman, R. K., Potter, M. A., McGlone, S. M., Cooley, P. C., Grefenstette, J. J., Zimmer, S. M., Wheaton, W. D., Quinn, S. C., Voorhees, R. E., & Burke, D. S. (2011). The benefits to all of ensuring equal and timely access to influenza vaccines in poor communities. *Health Aff (Millwood)., 30*, 1141–1150. https://doi.org/10.1377/hlthaff.2010.0778
32. Joseph, A., Madathil, K. C., Jafarifiroozabadi, R., Rogers, H., Mihandoust, S., Khasawneh, A., McNeese, N., Holmstedt, C., & McElligott, J. T. (2022). Communication and teamwork during telemedicine-enabled stroke care in an ambulance. *Human Factors, 64*, 21–41. https://doi.org/10.1177/0018720821995687

33. Epstein, J. M. (2002). Modeling civil violence: An agent-based computational approach. *Proceedings of the National Academy of Sciences of the United States of America, 99*(Suppl 3), 7243–7250. https://doi.org/10.1073/pnas.092080199
34. Yonas, M. A., Burke, J. G., Brown, S. T., Borrebach, J. D., Garland, R., Burke, D. S., & Grefenstette, J. J. (2013). Dynamic simulation of crime perpetration and reporting to examine community intervention strategies. *Health Education and Behavior, 40*, 87S-97S. https://doi.org/10.1177/1090198113493090
35. Tyshchuk, Y., & Wallace, W. A. (2018). Modeling human behavior on social media in response to significant events. *IEEE Transactions on Computational Social Systems, 5*, 444–457. https://doi.org/10.1109/TCSS.2018.2815786
36. Glenski, M., Weninger, T., & Volkova, S. (2018). Propagation from deceptive news sources who shares, how much, how evenly, and how quickly? *IEEE Transactions on Computational Social Systems, 5*, 1071–1082. https://doi.org/10.1109/TCSS.2018.2881071
37. AnyLogic: Simulation Modeling Software Tools & Solutions for Business. (n.d.). https://www.anylogic.com/ (accessed July 21, 2021).
38. Health Behavior: Theory, Research, and Practice—Google Books. (n.d.).
39. Henze, N. (1994). On Mardia's kurtosis test for multivariate normality. *Communications in Statistics—Theory and Methods., 23*, 1031–1045. https://doi.org/10.1080/03610929408831303
40. Sargent, R. G., & Balci, O. (2017). History of verification and validation of simulation models. In *2017 winter simulation conference* (*WSC*) (pp. 292–307). IEEE. https://doi.org/10.1109/WSC.2017.8247794
41. Kleijnen, J. P. C. (1995). Verification and validation of simulation models. *European Journal of Operational Research, 82*, 145–162. https://doi.org/10.1016/0377-2217(94)00016-6
42. Sargent, R. G. (2013). Verification and validation of simulation models. *Journal of Simulation, 7*, 12–24. https://doi.org/10.1057/jos.2012.20
43. Wu, Y., Stanton, B. F., Galbraith, J., Kaljee, L., Cottrell, L., Li, X., Harris, C. V., D'Alessandri, D., & Burns, J. M. (2003). Sustaining and broadening intervention impact: A longitudinal randomized trial of 3 adolescent risk reduction approaches. *Pediatrics, 111*, e32–e38.
44. Rosopa, P. J., Schaffer, M. M., & Schroeder, A. N. (2013). Managing heteroscedasticity in general linear models. *Psychological Methods, 18*, 335–351. https://doi.org/10.1037/a0032553
45. Ng, J. Y.-H., Yang, F., & Davis, L. S. (2015). Exploiting local features from deep networks for image retrieval. In *2015 IEEE conference on computer vision and pattern recognition workshops* (*CVPRW)* (pp. 53–61). IEEE. https://doi.org/10.1109/CVPRW.2015.7301272
46. Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). *Explaining and harnessing adversarial examples*. ArXiv.
47. Nguyen, A., Yosinski, J., & Clune, J. (2015). Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *2015 IEEE conference on computer vision and pattern recognition* (*CVPR*) (pp. 427–436). IEEE. https://doi.org/10.1109/CVPR.2015.7298640
48. P.M.C. Europe,—Abstract—Europe PMC. (n.d.).
49. Silverman, B. G., Hanrahan, N., Bharathy, G., Gordon, K., & Johnson, D. (2015). A systems approach to healthcare: Agent-based modeling, community mental health, and population well-being. *Artificial Intelligence in Medicine, 63*, 61–71. https://doi.org/10.1016/j.artmed.2014.08.006
50. Nikolov, D., Oliveira, D. F. M., Flammini, A., & Menczer, F. (2015). Measuring online social bubbles. *PeerJ Computer Science, 1*, e38. https://doi.org/10.7717/peerj-cs.38
51. APA PsycNet. (n.d.). https://psycnet.apa.org/buy/2009-09537-004 (accessed July 21, 2021).
52. Kastenmüller, A., Fischer, P., Jonas, E., Greitemeyer, T., Frey, D., Köppl, J., & Aydin, N. (2009). Selective exposure: The impact of framing information search instructions as gains and losses. *European Journal of Social Psychology*. https://doi.org/10.1002/ejsp.653
53. Bakshy, E., Messing, S., & Adamic, L. A. (2015). Political science. Exposure to ideologically diverse news and opinion on Facebook. *Science, 348*, 1130–1132. https://doi.org/10.1126/science.aaa1160
54. Safe and Effective Messaging and Reporting | Suicide Prevention Resource Center. (n.d.). https://sprc.org/keys-success/safe-messaging-reporting (accessed July 21, 2021).

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Amro Khasawneh[1] · Kapil Chalil Madathil[2,3] · Kevin M. Taaffe[3] · Heidi Zinzow[4] · Amal Ponathil[5] · Sreenath Chalil Madathil[6] · Siddhartha Nambiar[7] · Gaurav Nanda[8] · Patrick J. Rosopa[4]

1 Industrial Engineering Department, School of Engineering, Mercer University, Macon, GA, USA

2 Glenn Department of Civil Engineering, College of Engineering, Computing and Applied Sciences, Clemson University, Clemson, SC, USA

3 Industrial Engineering Department, College of Engineering, Computing and Applied Sciences, Clemson University, Clemson, SC, USA

4 Department of Psychology, College of Behavioral, Social and Health Sciences, Clemson University, Clemson, SC, USA

5 Glenn Department of Civil Engineering, College of Engineering, Computing and Applied Sciences, Clemson University, Clemson, SC, USA

6 Systems Science and Industrial Engineering Department, Thomas J. Watson College of Engineering and Applied Science, Binghamton University, Binghamton, NY, USA

7 North Carolina State University, Raleigh, NC, USA

8 Purdue University, West Lafayette, IN, USA