Understanding Virtual Epidemics:
Children's Folk Conceptions of a Computer Virus

Yasmin B. Kafai, University of California, Los Angeles, 2331 Moore Hall, kafai@gseis.ucla.edu

Abstract: The recent interest in virtual worlds for studying epidemics has created promising educational opportunities. Our work investigates the annual outbreak of Whypox, a virtual epidemic in Whyville.net, a virtual world with over 1.2 million registered players ages 8-16. We examined online and classroom participants' understanding of a computer virus using surveys and design activities. Our analyses reveal that students have a mostly naïve understanding of a computer virus influenced by mythological or anthropomorphic perspectives; only few were able to describe computational elements. The 35 students who participated in a curricular intervention in addition to the virtual epidemic shared these naïve conceptions initially, but developed more sophisticated views after the intervention. The discussion addresses possible explanations for students' conceptions and implications for future instructional designs.

It is only recently that the science community has shown an interest in virtual worlds as research laboratories for simulating human activities in events on a large scale. According to Bainbridge (2007), virtual worlds can be described as "complex physical spaces, where people can interact with each other and with virtual objects, and where people are represented by animated characters" that "require inhabitants to constantly experiment with unfamiliar alternatives, rationally calculate probable outcomes, and develop complex theoretical structures to understand their environment" (p. 472-475). These features have made them promising observatories for social and economic sciences (Castronova, 2005). The 2005 outbreak of a virulent epidemic in World of Warcraft suggested that virtual worlds could also be used to inform epidemiologists about the patterns of human behaviors and interactions that otherwise could not be studied for ethical reasons (Lofgren & Fefferman, 2007).

In this paper, we introduce the idea of using virtual epidemics as part of a K-12 science curriculum about infectious diseases (NRC, 1995). For our research, we used the annual outbreak of Whypox in Whyville.net, a virtual world with over 1.2 million registered players ages 8-16 years. Whypox is a virtual epidemic that infects players by having red pimples appear on their avatars and by interrupting their ability to chat with "sneezing" (i.e., typed words are replaced by "achoo") (Author et al., 2007). In prior studies, we have examined students' participation in the virtual epidemic as part of a sixth grade elementary science curriculum and their discussions about the Whypox epidemic in relation to natural infectious diseases (Neulight, Author, Kao, Foley & Galas, 2007). In this analysis, we realized that students also needed a better understanding of the underpinnings of the virtual epidemic itself beyond observing its behavioral impact. Like natural infectious diseases, virtual epidemics rely on a virus for transmission, in this case on a computer virus whose design features determine the disease vector, i.e., the mode, probability of infection, and its impact on the host.

While many students have heard, and even experienced, the impact of a computer virus such as 'Melissa', 'MyDoom', or 'Sasser' – names of past prominent computer virus infections, we know of no research to date that links these experienes to informal knowledge. The purpose of this paper is to map out the different dimensions of students' conceptions of a computer virus. We asked 285 online players in Whyville to describe their understanding of a computer virus and possible forms of protections against it. We also asked 35 sixth grade students to provide drawings and designs of a new Whypox virus. Finally, we examined whether or not the lived experience of a virtual epidemic would inform participants understanding of a computer virus. Our discussion addresses possible explanations for students'

To appear in the *Journal of Science Education and Technology*

conceptions, methodological issues, and implications for K-12 science education (NRC, 1995) and technology education (ISTE, 2000) to amplify possible connections between virtual and natural viruses.

Background

Our work builds on emerging research in the natural sciences where the simulation of behaviors has become a new method of scientific inquiry. Using this method, researchers exploit the similarities between viral distribution patterns in networks with those in the natural world. For instance, MacIsaac and Muirhead (2005) studied infestation patterns of fleas in lakes and use the spreading patterns of a computer virus to determine whether particular ecological contexts fostered faster infestation – which they confirmed in their analysis. The study of the viral plague in the virtual world of World of Warcraft uses this approach to understand the nature of human behaviors in large communities (Lofgren & Fefferman, 2007). Some have questioned how far these analogies can go, given the actual differences between real epidemics and virtual plagues, the fluidity of the natural environment, and a computer virus' common intent to attack a static number of machines (Boman & Johansson, 2007). This beginning set of studies illustrates the interdisciplinary nature of this method that brings together researchers from the fields of computer science, epidemiology and the natural sciences.

Previous studies in K-12 science and technology have used simulations on different technology platforms to help learners understand aspects of naturally occurring infections. For instance, Wilensky and Stroup (1999) had learners use a concurrent programming environment to help them understand the probability of contracting the disease whereas Collela (2000) used handheld devices in classroom to help students simulate and understand the spread of disease. Other approaches have recreated sanitary conditions in historic virtual worlds to guide students in their inquiry for causes for outbreaks (Dede et al., 2002). Most of these technology platforms also facilitated the repeated setup of virtual epidemics, thus allowing students to manipulate different parameters to better understand the impact on disease spread or probability of infection. These instructional uses of virtual epidemics have proven to be effective motivators in getting students interested in understanding different aspects of infectious diseases. However, all these studies used simulations for learning about natural diseases, and were not concerned with what students understood about the design of the computer virus while running the simulations.

Now there is also an interest to make this computational aspect of virtual epidemics accessible to students. We have already mentioned the emerging field in the natural sciences that uses virtual epidemics as an analytical tool to understand patterns of infestations and disease spread in the natural world. We think it's equally important that students understand how a computer virus functions given its prominent threat in the networked world. The destructive impact of a computer virus infection is felt as much by individual users as it is by large corporations whose communication networks might come to a standstill. While the purchase of protection software with ever vigilant updates can prevent such intrusions, few understand the mechanisms underlying a computer virus, worm, bot or trojan horse – to name but a few of the variants – created by hackers.

In absence of any prior studies on children's understanding of computer virus, we selected research on children's understanding of natural virus infections as a starting point for our investigation. Most of this research has examined children and youth's understanding within the context of AIDS (e.g., Sigelman et al, 1996). These studies offer some models of how young learners begin to develop concepts of germs, disease, spread and other aspects (Parmalee, 1992; Siegal, 1988; Siegal & Peterson, 1998; Solomon & Cassimatia, 1999). Of particular relevance to our research is Au and Romo's (1999) work on children's folk biology that examined different explanations that underlie children's reasons. They distinguish between behavioral explanations that focus on surface features without any causal explanations, mechanical explanations that focus on the movement of germs and the spread of disease, and finally biological explanations that take into account the reproductive elements of disease. Given the exploratory, nature of our research, we also requested explanations in visual format by asking students to draw a computer virus.

While questions about the nature of a computer virus are hypothetical, the participation in a Whypox outbreak in Whyville.net is experiential and provides players with a first hand experience over several days or weeks (Author et al., 2007). It should be noted that not all participants contract the disease but the consequences — the impact in online appearance and communicative interactions — of the outbreak are visible throughout the whole community. It is unclear to what extent such experiential participation can lead players to consider the underlying mechanics of what is happening. For that reason, we asked online users about their understanding of a computer virus before and after the virtual epidemic. In addition, we worked with two sixth-grade classes that participated in an infectious disease curriculum unit, which addressed the nature of computer viruses and their relationships to the virtual epidemic. We also asked those students to design a new version of the Whypox virus because unlike natural viruses that underlie evolutionary mutations, the designed nature of a computer virus invites such activities that could have instructional applications.

Research Approach

Participants

Our online sample consisted of 285 players who consented to participate in the research study's surveys posted before and after the outbreak of the virtual epidemic, called WhyPox. These users were recruited via postings and online town hall meetings on the site. In order to participate, users had to print out consent and assent forms and mail or fax them to the university. Our sample was representative of the larger Whyville population: 68% of participants were girls with a median age of 12.4 years (elementary: 6-11 years; n=90; middle school: 12-14 years; n=188; and high school: 15-18 years; n=27). We did not collect information about the ethnic background of online players. In addition, 35 students from two sixth-grade classes (ages 10-12) and their science teacher agreed to participate in the study. The students attended a laboratory school that is affiliated with a large, urban university and comprised a diverse ethnic sample representative of the California. In both classes, we had an equal group of boys and girls. Over 85% of these students have computer and Internet access from their homes (Kafai & Sutton, 2001). The classroom teacher had over twenty years of experience working in elementary schools and teaching science. In addition, a research team consisting of university faculty, a postdoctoral fellow, and graduate students documented classroom and online activities.

Settings

The virtual epidemic took place in a virtual world, called Whyville.net. At the time of the study in Winter 2005, Whyville had about 1.2 million registered users and could handle 4,000 concurrent users on the server. Whyville was designed with the intention to provide a space for children interested in science and thus features a large number of different science activities in addition to games (Author & Giang, in press). Players receive a virtual salary for each completed science activity with which they can purchase face parts, clothes and accessories designed by other Whyvillians for their online avatars. In addition, Whyville users can socialize in multiple ways by chatting publicly with each other or whispering privately to other members, sending y-mail and contributing to the Whyville Times newspaper. About once a year a virtual epidemic called Whypox is launched concurrently with the arrival of the flu season. Whyvillians who become infected with Whypox have red pimples appear on their avatars' faces and their chat activities are interrupted by sneezing, i.e., some words typed will be replaced by 'achoo'. All users can go to the Whyville's virtual Center for Disease Control (vCDC) where they can track the disease outbreak on a daily basis within the Whyville population, read about past cases of Whypox, and post predictions about causes and cures. In addition, Whyvillians can use tools that simulate outbreaks of diseases by manipulating variables such as the duration of a disease and number of initial infections.

The participation in Whyville and discussions about computer virus and Whypox were also integrated into an eight-week teacher-led curriculum about infectious diseases (see also Neulight et al., 2007). Some of the activities that students participated as part of their science curriculum included: watching videos about specific diseases and the nature of germs; examining cell structures under the microscope; doing hands-on experiments that simulated the spread of an infectious disease; completing worksheets about cells, bacteria, and viruses; and using online tools to research specific diseases. When Whypox hit Whyville during week five, the teacher facilitated whole-class discussions to address what was happening on Whyville. These discussions occurred approximately twice a week for about thirty minutes until the end of the study. As part of the classroom discussions, students were asked to create drawings and descriptions of a computer virus. The teacher and students also created a graph on a large piece of paper that displayed on one axis the number of Whypox infections in both classes and displayed on the other axis the date of infection. In addition to the discussions about Whypox, the teacher asked students to act as consultants to Whyville and design the next Whypox epidemic, i.e., to outline parameters for the next virtual epidemic. At week eight, students were administered the post-disease survey and a survey about Whypox.

Data Collection and Analyses

In the online survey, all 285 participants were asked to answer two questions: (1) "How would you describe to someone else what a computer virus is?" and (2) "What can you do to protect your computer against getting a computer virus?" This survey was given before the outbreak of Whypox and then again, eight weeks later, after the outbreak was over. The 35 classroom students also participated in this survey but were in addition given a blank sheet of paper with the directive: "Draw a picture of a computer virus – What do YOU think it looks like?" At the conclusion of the curricular unit, we asked the 35 students to design the next Whypox virus in the form of a proposal to the company that runs Whyville.net. Our data analysis is based on the 320 completed online surveys and 28 drawings and 27 proposals from classroom students. For all the answers, drawings and proposals, coding schemes were developed through an iterative process based on the responses received by students. The inter-rater agreement was established at 85%. The particular codes for each of the respective activities will be explained in the findings section together with the outcomes.

Findings

Understanding of a Computer Virus

Participants' understanding of a computer virus could be grouped into three major categories: behavioral, anthropomorphic, and computational. Over 50% of all answers to the question "How would you describe to someone else what a computer virus is?" were behavioral because the explanation focused on what the virus does to the computer distinguishing between unspecified ("bad things happen to your computer", or "it messes up your computer") and specific actions ("it deletes programs" or "it slows down the computer"). Twenty-three percent of the answers were placed into the anthropomorphism category because participants resorted to using human experience in getting an infectious disease as an explanation for describing a computer virus (e.g., "it's like getting the flu" or "it's like getting sick"). Only twelve percent of participants provided computational explanations. Here participants referred to the computer virus as a piece of code with instructions (e.g., "it's a program" or "it's a piece of code with instructions") or a reference to the man-made nature of computer viruses and their ability to replicate (e.g., "someone wrote the code" or " it replicates"). In addition, we had categories "Don't know" and "other" (see Table 1). In general, there were no significant differences pre/post survey ($p > .10$, z-test for dependent proportions) responses. Only significant differences were that middle school students provide significantly greater number of anthromorphic responses across time (23% v 36%) for Table 1, and for Table 2, middle score students provided significantly fewer avoidance responses across time (4% v 11%) and significantly fewer (17% v 7%) sourcing responses across time. Thus, only pre-survey results were presented in the tables.

Table 1: Definitions of a computer virus

	Elementary	Middle	High
Behavioral			
not specific	27%	26%	41%
specific	29%	28%	22%
Anthropomorphic	21%	23%	26%
Computational	8%	10%	4%
Other	12%	6%	4%
Don't know	3%	7%	4%
	n=90	n= 188	n=27

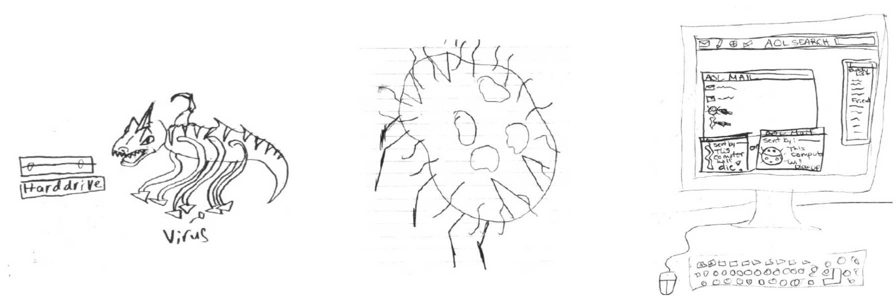
Table 2: Protections against a computer virus

	Elementary	Middle	High
Protection	58%	53%	66%
Avoidance	6%	4%	11%
Inactivity	18%	21%	17%
Sourcing	11%	17%	6%
Other	1%	3%	0%
Don't know	7%	3%	0%
	n=90	n= 188	n=27

A second question asked students "What can you do to protect your computer against getting a computer virus?" referring to protective actions against getting a computer virus (see Table 2). Over 56% of participants chose protection measures such as purchasing and installing commercial anti-virus software and setting up a firewall followed by more behavioral measures such as not opening email attachments (19%) or not going to particular websites (7%). Only 11% of participants considered sourcing such as checking the name of the sender. While for this question more than one answer was possible, only few participants were able to list more than one measure. The question about virus protection revealed that nearly half of all participants knew about anti-virus software but few knew or mentioned sourcing or other activities as viable strategies to ward off a virus infection.

Drawings of a Computer Virus

The 28 drawings of a computer virus were coded according to different categories. The first category of drawings represented various monsters or creatures with a humanoid body and strange facial features with claws for arms. Sometimes drawings featured a scary face with teeth and claws. Over 30% of all initial drawings choose such a representation for a computer virus but decreased to 17% after classroom readings and discussions about computer virus (see Figure 1). In a second category of drawings the computer virus resembled circular cells of biological bacteria or viruses under the microscope and was chosen by 25% of all students initially, a category that increased after the intervention to 50%. A third category of drawings illustrated computers with results from a virus attack.

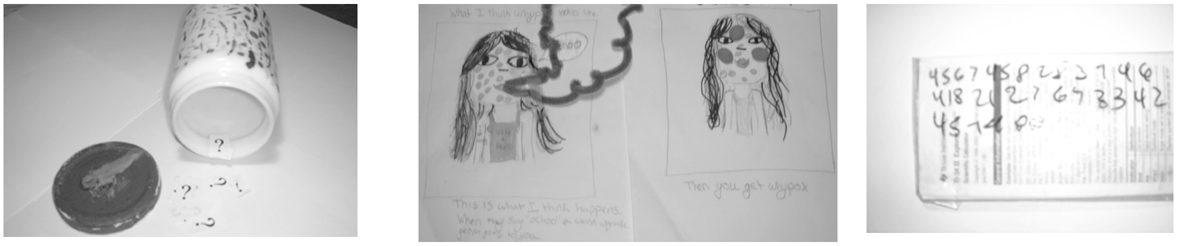


Drawing 1 Drawing 2 Drawing 3

Fig. 1: Drawings of a computer virus. In Drawing 1, the student drew the computer virus as a monster. In Drawing 2, the computer virus resembles a biological virus cell. In Drawing 3, students drew a computer screen with the results of a virus attack.

Designing the next Whpox

At the conclusion of the curricular unit, we asked students to design the next Whypox virus by writing a proposal to the company that owns Whyville. We received 27 proposals that outlined in great detail many aspects of the new Whypox virus (see Figure 2). The written proposals for Whypox2 (some contained drawings as well) were classified into four different categories, the first one being a replication of the original Whypox virus. Here students would describe and explain the symptoms of Whypox2. In some cases students would include explanations of how to treat or cure the virus and suggest the building of hospital facilities or the development of vaccine. Four students replicated the original WhyPox design with little or no variations.



Drawing 1 Drawing 2 Drawing 3

Fig. 2: Drawings of the Whypox2 virus. Drawing 1 shows a jar with Whyville viruses. Drawing 2 illustrates two different states of Whypox, one with how the virus travels through the air and one with Whypox infection. Drawing 3 shows a students' idea of computer code.

A second category of proposals would create a new disease with different symptoms, a different incubation period and/or different ways of spreading it. For example, Whyvillians would cough instead of sneeze or turn green instead of red or turn into green plants. Fourteen proposals (52%) created a new disease with new visible signs or mechanisms of infections. Two of the proposals added some additional elements such as variations in lengths or pimple design. A third category recognized Whypox2 as a learning tool and created games in which the player had to find a cure by passing through an obstacle course or earn more clams by playing science games to buy an expensive vaccine. Five proposals drew on the learning potential of the WhyPox virus. The final category described a new disease by explaining the computer programming of the virus. For instance, in one proposal the author described how the code would be passed from one person to the next while in others students generated what appears to be their idea of computer code. Only two proposals included computational elements by using program code or instructions in natural language that emulated program code.

Participation in Whypox

Finally, we investigated the impact of the Whypox participation by comparing the survey answers from the 35 classroom participants to a matched age sample (n=153) from the online participants. We found that classroom students not only became more specific but also developed more answers with computational elements when defining a computer virus (see Table 3).

Table 3: Definitions of a Computer Virus

	Classroom		Online	
	Pre	Post	Pre	Post
Behavioral				
not specific	17%	9%	28%	35%
specific	23%	31%	29%	24%
Anthropomorphic	23%	17%	23%	27%
Computational	9%	40%	10%	6%
Other	6%	3%	6%	3%
Don't Know	23%	0%	4%	5%
	n=35		n=153	

Table 4: Protections against a Computer Virus

	Classroom		Online	
	Pre	Post	Pre	Post
Protection	54%	60%	52%	60%
Avoidance	12%	18%	7%	10%
Inactivity	12%	10%	20%	24%
Sourcing	5%	19%	18%	4%
Other	10%	0%	1%	2%
Don't Know	7%	3%	2%	1%
	n=35		n=153	

Though there generally no differences between pre/post survey response for both Classroom and Online participants, result showed that Classroom students provided significantly more computational responses and significantly fewer 'don't know' response across time ($p < .05$, z-test for dependent proportions).

When we assessed how many different strategies classroom students knew to protect against a computer virus, a major difference was in listing sourcing strategies (see Table 4). Participants from the classroom project also provide significantly more ways to protect themselves against a virus (i.e., sourcing) across time, while online participants provided significant significantly fewer sourcing responses (both $p < .05$, z-test for dependent proportions)

Discussion

Our findings illustrate that the large majority of online users have little understanding of a computer virus. Most of the provided explanations focus on behavioral aspects and not on the underlying structure or programs of a computer virus. We have some evidence that students transferred their understanding of the impact of natural viruses to computer viruses via the large number of anthropomorphism responses. Many students and online users provided descriptions such as "the computer has the flu" and thus making reference to the impact caused by the flu virus on the human body's functioning and well-being. Anthropomorphism is often displayed when explaining the nature of animal behavior in terms of human intentions and structures. A well-known example is in the naming of the "queen" bee assuming that this bee has a higher status role in the hive while in fact she is just following a different set of rules. In our case, the computer was seen as a human catching the flu, or the computer virus, thus explaining the loss of functioning or data.

We also need to be aware that there are other facets to understanding virtual epidemics, if we want to use them as learning environments in K-12 education. For one, the behavioral ramifications in virtual epidemics deserve a closer look. After all, this is why medical researchers are so interested in using virtual worlds for studying the epidemiology of infectious diseases. Simulations have been around for many years but these have not been able to factor in human interactions and behaviors. Virtual worlds that are populated by living being's avatars offer the opportunity, albeit with some reservations, of witnessing interactions during a real time epidemic outbreak. Thus, it is equally important to get an insight of what Whyville players understand about the possible ostracism or irresponsible behavior that might become apparent during a virtual epidemic and what this says about their understanding about the causes and incubation periods of infectious diseases.

These findings offer a first window into students' conceptual understanding of a computer virus but we need to be mindful of the methodological limitations. While understanding and creating visual representations are important facets of science learning, they do not work well for all concepts. These approaches seem to work well when we ask students to generate drawings of biological specimen such as a natural virus. But computer viruses really have no visual equivalent; theirs is more a written or symbolic representation in form of computer programming language. It could be possible that the request to draw a computer virus might have misled many students to seek recourse to the mythological or anthropomorphic representations, given their documented lack of programming knowledge. Even the representations generated by the classroom students, such as number code in drawing 3 in Figure 2, are at best surface representation of code when students student take the notion of computing too literal. A few were able to provide a list of instruction in natural language that in fact resembled what actual program code would look like. A lack of programming experience, very common in schools, provides the best explanation for the lack of explanations in this category. We see here clear implications for the connection between computation and inquiry, as it is already apparent in many science fields, but missing in today's K-12 technology and science education. Computer and natural viruses populate our worlds, real and virtual alike, with equally devastating repercussions. The inclusion of computer virus in the repertoire of infectious disease curriculum could build a bridge between technology and science. In fact, the understanding of one can be beneficial for the understanding of the other, and vice versa. This mutual beneficial relationship is worth further investigation.

References

Au, T. K., & Romo, L. F. (1999). Mechanical causality in children's "folkbiology". In D. Medin & S. Atran (Eds.), *Folkbiology*. Cambridge, MA: The MIT Press.

Bainbridge, W. S. (2007). The research potential of virtual worlds. *Science, 317*, 472-476.

Boman, M. & Johansson, S. J. (2007). Modeling epidemic spread in synthetic populations—Virtual plagues in massively multiplayer online games. In Akira Baba (Ed.), *Situated Play: Proceedings of the Third International Conference of the Digital Games Research Association (DiGRA)* (pp. 324-332). Tokyo, Japan: The University of Tokyo.

Castronova, E. (2005). *Synthetic worlds: The business and pleasure of gaming*. Chicago: Chicago University Press.

Colella, V. (2000). Participatory simulations: Building collaborative understanding through immersive dynamic modeling. *Journal of the Learning Sciences, 9*(4), 471-500.

Dede, C., Nelson, B., Ketelhut, D. J., Clarke, J., & Bowman, C. (2004). Design-based research strategies for studying situated learning in a multi-user virtual environment. In Y. B. Xxxx, W. A. Sandoval, N. Enyedy, A. S. Nixon & F. Herrera (Eds.), *Proceedings of the Sixth International Conference of the Learning Sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Kafai, Y. B., Feldon, D., Fields, D., Giang, M., & Quintero, M. (2007). Life in the Times of Whypox: A Virtual Epidemic as a Community Event. In C. Steinfeld, B. Pentland, M. Ackermann, & N. Contractor (Eds.), *Proceedings of the Third International Conference on Communities and Technology* (pp. 171-190). New York: Springer.

Kafai, Y. B. & Giang, M. (in press). Virtual Playgrounds. In T. Willoughby & E. Wood (Eds), *Children's Learning in a Digital World*. (pp. 196-217). Oxford: Blackwell Publishing.

Lofgren, E. T. & Fefferman, N. H. (2007). The untapped potential of virtual game worlds to shed light on real world epidemics. *Lancet Infectious Diseases, 7*, 625-629.

MacIsaac, J. R. & Muirhead, H. J. (2005). Development of inland lakes as hubs in an invasion network. *Journal of Applied Ecology, 42*(1), 80-98.

National Research Council (1995). *National science education standards*. Washington, DC: National Research Council.

Neulight, N., Kafai, Y. B., Kao, L., Foley, B., & Galas, C. (2007). Children’s Learning about Infectious Disease through Participation in a Virtual Epidemic. *Journal of Science Education and Technology*.

Parmelee, A. H. (1992). Wellness, illness, health, and disease concepts. In E. J. Susman, L. V. Feagans & W. J. Ray (Eds.), *Emotion, cognition, health, and development in children and adolescents* (pp. 155-164). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Siegal, M. (1988). Children's understanding of contagion and contamination as causes of illness. *Child Development, 59*(4-6), 1353-1359.

Sigelman, C., Estrada, A. L., Derenowski, E., & Woods, T. (1996). Intuitive theories of human immunodeficiency virus transmission. *Journal of Pediatric Psychology, 21*(4), 555-572.

Solomon, G. E., & Cassimatis, N. L. (1999). On facts and conceptual systems: Young children's integration of their understandings of germs and contagion. *Developmental Psychology, 35*, 113-126.

Wilensky, U., & Stroup, W. (1999). *Learning through participatory simulations: Network-based design for systems learning in classrooms computer supported collaborative learning*. Paper presented at the CSCL, Stanford University.

Acknowledgments

The work reported in this paper is supported by a grant of the National Science Foundation. The views expressed are those of the author and do not necessarily represent the views of the supporting funding agency or the University of California, Los Angeles.