UNIVERSITY OF CALIFORNIA, IRVINE

Assessing Understanding of Complex Causal Networks Using an Interactive Game

DISSERTATION

submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in Information and Computer Science

by

JOEL ROSS

Dissertation Committee:
Professor Bill Tomlinson, Chair
Professor Paul Dourish
Professor Bonnie Nardi

2013

Table of Contents

List of Figures

Page

List of Tables

Acknowledgments

The work presented in this dissertation was generously supported by the National Science Foundation Grant ISS-0644415, by a Research Grant from the Alfred P. Sloan Foundation, by the Donald Bren School of Information and Computer Sciences, and by the California Institute of Telecommunications and Information Technology (Calit2).

Collaboration is at the soul of all research, and this dissertation would not have been possible without the help and support of countless others. To all of you who have touched this work in some way, thank you.

Thanks to my advisor, Bill Tomlinson. For the past five years, you have backed my efforts—both success and failures. You provided a welcoming home for me to investigate new topics, while constantly urging me to reach further. Your support has been the foundation, the bedrock on which I have been able to steadily build while trusting that the ground will not fall out from under me.

Thanks to my dissertation committee, Paul Dourish and Bonnie Nardi, who kept me grounded with your questions and challenges to my assumptions. You continue to teach me how to think through my work and latch onto the successful contributions.

Thanks to all those who helped worked on the Causlings game: Alicia Chan, Daniel Lobato Garcia, Raminder Goraya, Sahand Nayebaziz, Duyet Nguyen, Jack Pan, Dinorah Rodriguez, Anna Slykhous, and Shelly Yu. Your coding, design, testing, and input enabled me to rapidly complete and deploy this project and time got tight.

Special thanks to Eugenia Gabrielova, for your enthusiasm, for always having time to talk, and for the plethora of shared animal pictures. And special thanks to Ankita Raturi, for being a friendly face and a voice of commiseration during the long haul, and for making sure that I didn't have to write alone.

Thanks to all the members of the Social Code Group, past and present. Thanks to Eric Baumer, who showed me the way. Thanks to Six Silberman, who gave me ideas. Thanks to Derek Lyons, who laid the foundation. And thanks to all those who worked with me across years of projects, laying the basis for this work: Nadine Amsel, Robert Beckman, Damon Domjan, Ellen Eramya, Josh Fulmer, Oliver Holmes, Michael Riccobono, Nitin Shantharam, Robert Simpson, Andrew Zaldivar, and all the other members of our lab. You made the lab my academic home away from home—an environment in which I could thrive.

Thanks to Professors Rebecca Black and Melissa Mazmanian, for your guidance and experience. Also thanks to my fellow graduate students Jed Brubaker, Marisa Cohn, Lynn Dombrowski, Ellie Harmon, Lilly Irani, and Nick Seaver, and the members of the CHI 2011 Doctoral Consortium, all who listened and gave me advice on my research, no matter what direction it was taking at the time. All your insights and guidance helped steer me through some of the darkest waters.

Also thanks to Chris O'Neal and the 2011 Pedagogical Fellows Program, for giving me the chance to practice what I love and motivating me to complete this work. And special thanks to Dan Flynn, for the countless hours of collaboration and for understanding when the dissertation had to come first.

Thanks to my close friends Robert Simpson, Jeff Johnston, Peter Thompson, Rosie Friedland, Amanda Johnston, and Ellen Kern, for gaming with me and having fun, despite my distraction and lack of time. You put up with all my aborted efforts while I saw this work through to the end. I truly treasure our friendship and all the time and trips we've spent together.

Thanks to my family: Dottie, Marilynn and Arnold, Debbie and Matt, and Bob and Sally, for all the warm meals and warm company.

Thanks to my parents, for pretending to understand what I was doing and continuing to nag me, no matter how far away I was.

And most importantly, thanks to my beautiful wife Peggy. For giving me permission to work, and also permission to relax. For putting up with my changing moods, and encouraging me through the highs and the lows. For listening to all my rants and complaints, no matter their importance. For all of the little tasks and favors, when I was too busy or preoccupied. And for being a beautiful, smiling face at the end of the day—one that I love to see, and always will. You are my second half, my closest friend, and what is best in life.

בָּרוּךְ אַתָּה יְיָ אֱלֹהֵינוּ מֶלֶךְ הָעוֹלָם שֶׁהֶחֱיָנוּ וְקִיְּמָנוּ וְהִגִּיעָנוּ לַזְּמַן הַזֶּה:

Curriculum Vitae

ACADEMIC APPOINTMENTS

UNIVERSITY OF PUGET SOUND

- Visiting Assistant Professor of Computer Science, 2012 - present

UNIVERSITY OF CALIFORNIA, IRVINE

- Graduate Research Assistant, 2007 - 2012
- Teaching Associate (Instructor), 2011
- Teaching Assistant, 2009-2012

EDUCATION

Ph.D. Information and Computer Science; 2013
University of California, Irvine

- Dissertation: *Assessing Understanding of Complex Causal Networks Using an Interactive Game*
- Advisor: Bill Tomlinson

M.S. Information and Computer Science, Informatics (Ubiquitous Computing); 2009
University of California, Irvine

- Advisor: Bill Tomlinson

B. A. *magna cum laude*, Double Major: Mathematics (Computer Science), English; 2006
The Colorado College, Colorado Springs, CO

- Distinction in Mathematics (emphasis on Computer Science)
- Distinction in English
- English Thesis: *A Conflict of Views: Tensions in the Multi-Plot Narrative*

PUBLICATIONS

JOURNAL PAPERS

Ross, J., Simpson, R., and Tomlinson, B. (2011). "Media Richness, Interactivity, and Retargeting to Mobile Devices: A Survey". In: *International Journal of Arts and Technology*, 4(4), 442-459. Inderscience.

Ross, J. and Tomlinson, B. (2011). "Negabehaviors and Environmental Sustainability". In: *Journal of Sustainability Education*, 2(1). Online.

Tomlinson, B., Yau, M. L., Baumer, E., **Ross, J.**, Correa, A., & Ji, G. (2009). "Richly Connected Systems and Multi-Device Worlds". In: *Presence: Teleoperators & Virtual Environments*, 18(1), 54-71. MIT Press.

PEER-REVIEWED FULL CONFERENCE PROCEEDINGS

Ross, J., Shantharam, N., and Tomlinson, B. (2010). "Collaborative Filtering and Carbon Footprint Calculation". In: *Proceedings of the 2010 International Symposium on Sustainable Systems and Technology (ISSST)*. Washington D.C.: IEEE.

PEER-REVIEWED SHORT CONFERENCE PROCEEDINGS AND WORKSHOP PAPERS

Tomlinson, B., **Ross, J.**, Andre, P., Baumer, E., Patterson, D., Corneli, J., et al. (2012). "Massively Distributed Authorship of Academic Papers". In: *alt.chi session of CHI 2012 extended abstracts on human factors in computing systems*. (11-20). Austin, Texas, USA: ACM. [47% acceptance rate]

Ross, J. (2011). "Pervasive Negabehavior Games for Environmental Sustainability". In: *Doctoral Consortium session of CHI 2011 extended abstracts on human factors in computing systems.* (1085-1088). Vancouver, Canada: ACM. [32% acceptance rate]

Ross, J. and Tomlinson, B. (2010). "How Games Can Redirect Humanity's Cognitive Surplus for Social Good". In: *ACM Computers in Entertainment (CIE)*, 8(4). ACM.

Silberman, M.S., Irani, L., **Ross, J.** (2010). "Ethics and Tactics of Professional Crowdwork". In: *XRDS: Crossroads, The ACM Magazine for Students*, 17(2). ACM.

Silberman, M.S., **Ross, J.**, Irani, L., and Tomlinson, B. (2010). "Sellers' problems in human computation markets". In: *Proceedings of the ACM SIGKDD Workshop on Human Computation (HCOMP '10)*. (18-21). Washington D.C.: ACM. [28% acceptance rate]

Ross, J., Irani, L., Silberman, M.S., Zaldivar, A., and Tomlinson, B. (2010). "Who are the Crowdworkers?: Shifting Demographics in Mechanical Turk". In: *alt.chi session of CHI 2010 extended abstracts on human factors in computing systems*. (2863-2872). Atlanta, Georgia, USA: ACM. [42% acceptance rate]

TECHNICAL REPORTS

Ross, J., Holmes, O., and Tomlinson, B. (2012). "Playing with Genre: User-Generated Game Design in LittleBigPlanet 2". LUCI Report 2012-003.

Ross, J., Amsel, N., Beckman, R., and Tomlinson, B. (2010). "EcoPath: Adding Spatial, Social, and Gaming Contexts to Personal Tracking Systems". Social Code Report 2010-01.

Ross, J., Zaldivar, A., Irani, L., and Tomlinson, B. (2009). "Who are the Turkers? Worker Demographics in Amazon Mechanical Turk". Social Code Report 2009-01.

TEACHING

COURSES TAUGHT

- Introduction to Computer Science, Puget Sound, Fall 2012
- Software Engineering, Puget Sound, Fall 2012

- Critical Writing in Information and Computer Sciences, UCI, Summer 2011
- TA: Informatics Core II *[Intro to programming in Python*], UCI, Winter 2012
- TA: Critical Writing in Information and Computer Sciences, UCI, Summer 2011
- TA: Game Technologies and Interactive Media, UCI, Spring 2011
- TA: Critical Writing, UCI, Summer 2010
- TA: Organizational Information Systems, UCI, Winter 2010
- TA: Social Analysis of Computerization, UCI, Fall 2009

GUEST LECTURER

- "Serious Games and Gamification". Computer Games and Society, UCI, Fall 2011.
- "Pervasive and Mobile Game Technologies". Game Technologies and Interactive Media, UCI, Spring 2011.
- "Persuading Using Pervasive Games". Computer Games and Society, UCI, Fall 2010.

PEDAGOGICAL DEVELOPMENT

- Pedagogical Fellow, UCI Teaching Learning and Technology Center, 2011 - 2012
- "Teach Smarter with EEE". UCI Teaching Learning and Technology Center Workshop Series, 2012. With Dan Flynn. Designed and conducted workshop on learning management system.
- TA Professional Development Program (TAPDP), UCI, 2011. With Dan Flynn. Designed and conducted two-day training workshop for 43 new TAs.

MENTORING & COLLABORATION WITH UNDERGRADUATES

- Damon Domjan, Summer 2010 - Fall 2011
- Oliver Holmes, Spring 2011 - Summer 2011
  (Calit2 SURF-IT Fellow, "Understanding Player Dynamics in Thwarting Games", 2011)
- Josh Fulmer, Summer 2010 - Spring 2011
- Bob Beckman, Summer 2009 - Spring 2010
- Nitin Shantharam, Winter 2008 - Spring 2010
  (Honors Thesis, "Collaborative Filtering and Carbon Footprint Calculation", 2010)
- Andrew Zaldivar, Fall 2008 - Spring 2009
- Robert Simpson, Summer 2008 - Winter 2009
  (Calit2 SURF-IT Fellow, "Human Mediated Networking", 2008)
- Michael Riccobono, Summer 2008
  (NSF REU, "Green IT", 2008)
- Ellen Eramya, Summer 2007 - Spring 2008

HONORS AND AWARDS

HONORS

- UCI Most Promising Future Faculty Dissertation Fellowship, 2012
- NSF Graduate Research Fellowship Honorable Mention, 2009
- UCI Public Impact Fellowship Honorable Mention, 2009
- Phi Beta Kappa Honor Society, 2006
- Frank Krutzke Senior Project Award in English, Colorado College, 2006

AWARDS

- Paul Butterworth Student Award ($1200, "The Better Carbon Calculator"), 2010. With Nitin Shantharam.
- Paul Butterworth Student Award ($4200, "EcoPath"), 2010. With Nadine Amsel.

SCHOLARSHIPS

- Dean's Fellowship, Donald Bren School of Information and Computer Sciences, 2007 - 2011
- National Merit Scholarship, 2002 - 2006

PRESENTATIONS

"Persuading 'Negabehaviors' with Pervasive Games and Stories". *Participant, NSF Persuasive Experiences Workshop*. Sept 23-26, 2010. Burbank, California.

"Harnessing the Wisdom of Crowds for Human Computation". *Boaster, Human Computer Interaction Consortium (HCIC)*. Feb 24-28, 2010. Fraser, Colorado.

"Globalizing Cognitive Labor: The Changing Demographics of Amazon Mechanical Turk". *Advancement to candidacy talk*. Dec 9, 2009. University of California, Irvine.

"Social Green-IT". *Guest Speaker, Osher Lifelong Learning Institute.* Nov 5, 2008. University of California, Irvine.

PRESS COVERAGE

"Can Computer Code Be Greener? Facebook Thinks So", D. Zax, *FastCompany.com*, 11/12/2010

"Mechanical Turk lets you make a few bucks online", N. Diakopoulos, *The Sacramento Bee*, 7/28/2009

SERVICE

FACULTY ADVISOR

- University of Puget Sound ACM Student Chapter, 2012 - present

CONFERENCE STUDENT VOLUNTEER

- Student Volunteer for ACM SIGCHI Conference on Human Factors in Computing Systems (CHI), 2012

REVIEWER

- ACM SIGCHI Conference on Human Factors in Computing Systems (CHI), 2008, 2010 - 2013
- ACM Technical Symposium on Computer Science Education (SIGCSE), 2012 - 2013
- ACM Designing Interaction Systems Conference (DIS), 2010, 2012

- ACM Conference on Computer Supported Cooperative Work (CSCW), 2010, 2012
- ACM Computers in Entertainment (CIE), 2010
- ACM Transactions on Computer-Human Interaction (TOCHI), 2010
- IEEE IT Professional Magazine, 2010
- IEEE Computer Graphics and Applications Magazine, 2007

MEMBERSHIPS

- Association for Computing Machinery (ACM), 2010 - present
- ACM Special Interest Group on Computer Science Education (SIGCSE), 2011 - present

SKILLS

COMPUTER SKILLS

- Java, Python, Ruby, C++, PHP, MySQL, HTML/CSS, Javascript
- Android platform, Ruby on Rails framework
- OpenGL, Java3D
- Microsoft Windows, Mac OS X, and Linux operating systems
- Blender, Final Cut Pro
- Advanced Microsoft Office (Word, Excel, Access, Front Page, PowerPoint)
- LaTeX
- R (basic)

EMPLOYMENT

Mathematics and Computer Science Department, University of Puget Sound, Tacoma, WA,
Visiting Assistant Professor 2012 - present

- Taught *Introduction to Computer Science* and *Software Engineering*

Informatics Department, UC Irvine, Irvine, CA, 2007 - 2012
Graduate Student Researcher

- Research projects include: EcoRaft, GreenScanner, Peer-to-Peer Sensor Sharing, studies of Amazon Mechanical Turk, retargeting media to mobile devices, Better Carbon, EcoPath/EcoDefender, Negabehaviors, Thwarting Games, User-Generated Game Design, Causality Project (visualization), *Causlings*

Summer Session, UC Irvine, CA, 2011
Teaching Associate

- Taught *Critical Writing in Information and Computer Sciences*

Economics and Business Department, Colorado College, 2006 - 2007
Paraprofessional

Abstract of the Dissertation

Assessing Understanding of Complex Causal Networks Using an Interactive Game

By

Joel Ross

Doctor of Philosophy in Information and Computer Science

University of California, Irvine, 2013

Professor Bill Tomlinson, Chair

Assessing people's understanding of the causal relationships found in large-scale complex systems may be necessary for addressing many critical social concerns, such as environmental sustainability. Existing methods for assessing systems thinking and causal understanding frequently use the technique of cognitive causal mapping. However, the logistics of this methodology may miss valuable and informative indicators of reductionist and linear thinking, both of which conflict with systems understanding.

This dissertation explores how interactive computer systems can aid in the assessment of causal understanding, allowing educators to perform more in-depth analysis of how subjects engage with the process of causal mapping. In addition, it considers how computer games as a particular form of interactive system may be able to support assessment. Games are framed as effectively supporting learning and education and although assessment is a key component of education, the use of video games for performing assessment is under-explored.

To address these topics, I present a prototype interactive game system based on Plate's (2006) framework for assessing causal understanding through cognitive causal mapping. I tested this prototype in a user study with both student and non-student subjects. Through this study, I found that evaluating the structural forms of causal maps created in an interactive system can suggest

the presence of reductionist thinking, while the sequence of causal map construction can indicate the presence of linear thinking. Furthermore, I found that although games as interactive systems can be effective in enabling learning, they may be less readily effective in supporting stand-alone assessments due to requiring an *a priori* understanding of the complex game system used in assessment, as well as traditional educational assessment contexts not supporting the forms of feedback critical to game-based learning.

These results indicate how the linear narratives prominently found in both education and games may interfere with effective systems thinking. This dissertation thus suggests that educators in both formal and informal education contexts should consider alternative, non-narrative curricula and games for teaching and assessing causal understanding of complex systems.

Chapter 1. Introduction

As humans we sometimes struggle with understanding cause and effect in the world around us. We are good at figuring out causal relationships for physical entities (e.g., if we kick a can, it goes flying as an effect), or when cause and effect are close to each other in time and space (e.g., if we touch the hot stove, we quickly get burned). But causal relationships that are not directly apparent, or have a greater distance between cause and effect, confound us. We can't always identify the causes that can explain what will make our flower gardens to grow, why our computers won't connect to the Internet, or how the market price of oil impacts the natural environment. In these cases, the link between cause and effect may be too obscured or too complex for us to understand readily.

A great number of the causal relationships we would like to understand are difficult to comprehend because they are part of larger complex systems. For example, flower gardens are part of a broader ecosystem that includes components such as the soil, the weather, other plants and weeds, any resident insects, a human with a trowel, and a dog that just won't stay out of the petunias. This ecosystem is a *complex system* (see e.g., Hmelo-Silver and Azevedo, 2006; Jacobson and Wilensky, 2006) as it has multiple heterogeneous components, all of which interact with one another to produce an emergent behavior that is not obvious from just considering the individual parts. Understanding these complex systems requires a form of cognition that is often called *systems thinking* (Booth Sweeney and Sterman, 2000; Doyle, 1997), and is in many ways the opposite of the reductionist thinking commonly taught in schools (Plate, 2006). Rather than understanding a system by considering and then linearly combing its constituent parts, systems thinking proposes a view in which "events and actions have multiple causes and consequences, and where order and structure coexist at many different scales of time, space, and organization" (Jacobson and Wilensky, 2006). Systems thinking requires skills in recognizing the existence of feedback processes, interactions between stocks and flows, the impacts of delayed effects, and the presence of multiple non-linear cause-and-effect relationships.

Systems thinking is increasingly seen as critical to comprehending the complexities of the modern world. Systems ecologist H.T. Odum wrote that "if the bewildering complexity of human knowledge developed in the twentieth century is to be retained and well used, unifying concepts are needed to consolidate the understanding of systems of many kinds and to simplify the teaching of general principles" (Odum, 1983). Many researchers and educators have adopted this view, and systems thinking has gained a prominent position in current education standards and research. For example, the U.S. National Research Council includes "Systems, order, and organization" as one of its five "unifying concepts and processes" for science education (National Research Council, 1996). Similarly, systems thinking is included as a key 21st Century Skill, with the suggestion that students must learn to "analyze how parts of a whole interact with each other to produce overall outcomes in complex systems" (Partnership for 21st Century Skills, 2001). These types of curriculum standards have led to the creation of a variety of activities, interventions, and systems designed to increase people's ability to think about systems, both in academia and in industry (e.g., Senge and Sterman, 1992).

Understanding systems and their causal relationships is especially important when considering issues in environmental sustainability. Achieving environmental sustainability is one of the most pressing concerns facing the whole of the human race. This challenge includes a wide range of issues, from fossil fuel pollution (IPCC, 2007), to anthropogenic climate change (IPCC, 2007; Pew Center on Global Climate Change, 2009), to biodiversity loss (Thomas et al., 2004; Wilson, 1999). These issues are the results—that is, "effects"—of incredibly complex systems that include the behaviors of countless heterogeneous actors, which moreover are distributed spatially across the globe and unfold over years or decades. Effectively dealing with the vast scope of these issues thus required thinking about them as systems. For example, Sterman and Booth Sweeney explain how the systems thinking concepts of stocks and flows relate to CO_2 emissions and global climate change:

> "The higher the concentration of greenhouse gases in the atmosphere, the higher the global temperature will eventually become. The stock and flow structure of the global climate means stabilizing emissions near current rates will not stabilize the climate, but ensures continued growth in GHG [greenhouse gas] concentrations, and continued warming... it is as simple as filling a tub: Humanity is injecting CO_2 into the atmosphere at about twice the rate it is drained out" (Sterman and Booth Sweeney, 2002).

By applying systems thinking to the environmental issue of climate change, Sterman and Booth Sweeney point how (for example) the action of "reducing carbon emissions by 10%" will *not* have the effect of reducing global warming—only slowing the rate at which climate temperatures increase. While a systems perspective may make the process "as simple as filling a tub," humans struggle to understand the cause-and-effect relationships of their actions because of the distributed complexity of environmental issues.

As humans we may not be readily equipped with the cognitive skills to think intuitively about complex systems on the spatial and temporal scale of environmental issues (Tomlinson, 2010). The human mind has been shaped by evolution to have particular strengths and weakness (Gigerenzer, 2000; Gigerenzer and Todd, 1999; Lyons and Santos, 2006; Tomlinson, 2010), and as such humans are sensitive to certain types of cause-and-effect relationships. For example, humans most intuitively understand causal relationships that involve direct, linear, and instantaneous connections between cause and effect, that have causes spatially and temporally localized to the person, and that are driven by a centralized agency (rather than being "side effects" of other actions) (Grotzer and Perkins, 2000). Thus people intuitively understand that kicked cans move and hot stoves burn, but have difficulties comprehending how electricity flows in a circuit or how components interact in an ecosystem.

A number of studies have shown how people struggle to understand the multiple, non-linear causal relationships that make up complex systems such as those surrounding environmental sustainability. For example, Raia (2005) found that undergraduates "tend to conceptualize dynamic systems in static disjointed terms, considering the isolated behavior of the constituent components," and "identify a single causal force, or linear chain of unique causal forces to explain complex natural phenomena." Students did not identify effects as the emergent results of multiple causes, instead attributing effects to a single centralized cause. In this way students engage in linear thinking (considering a system as a linear sequence of events) rather than systems thinking. Sterman and Booth Sweeney show that even highly educated graduate students such as those from Harvard and MIT fail to understand the dynamic relationships in both simple systems like a filling bathtub (Booth Sweeney and Sterman, 2000) and complex systems such global temperature levels (Sterman and Booth Sweeney, 2002). Similarly, Moxnes found that professionals often failed to properly understand non-linear cause-and-effect in complex systems, even in contexts related to their work (Moxnes, 2000). In a study where 59 out of 82 participants were professionals in the fishing industry, the median participant overfished a simulated ecosystem 52% of the time; in another study, participants looked for causes outside the boundaries of the simulation in attempts to find direct, linear causal relationships.

These examples all demonstrate how people can often fail to understand causal relationships in complex systems. The founder of system dynamics, Jay W. Forrester, wrote that "Our intuitive judgment is unreliable about how these systems will change with time, even when we have good knowledge of the individual parts of the system" (Forrester, 1961). Furthermore, people may not even be aware of their own misunderstandings: players of the MIT Beer Game (a game simulating a complex system, described in detail in Chapter 3) often initially conclude that the game is "rigged," rather than realizing that they have a flawed understanding of the system (Goodwin and Franklin, 1994). When dealing with the non-linear, emergent cause-and-effect relationships that make up complex systems, people are often not even aware of what they don't understand. Identifying the flaws in our intuition and understanding is of crucial importance when dealing with complex systems. This assessment is particularly important when attempting to address issues in environmental sustainability, either through individual behavior changes or by instituting new public policies.

Consider for example the history of ocean iron fertilization (OIF). In OIF, large deposits of iron are inputted into the oceans in order to stimulate the growth of phytoplankton algae. Iron usually acts as the limiting nutrient for algae growth—however, by added extra iron, larger algae blooms can be produced. The increased amount of algae leads to a greater amount of photosynthesis, pulling more CO_2 from the atmosphere and into the ocean, where it is stored. As Strong, Cullen, and Chisholm explain:

> "Phytoplankton in the lighted surface layer take up nutrients (e.g., nitrate and phosphate) and grow, converting CO_2 to organic matter that fuels marine food webs. Some of the organic matter—for example, senescent phytoplankton, fecal pellets, and aggregated debris—sinks to the deep ocean where it decomposes, releasing CO_2 and nutrients while consuming oxygen. When the ocean carbon cycle is roughly in balance, this carbon- and nutrient-rich deep water does not reach the surface for decades to hundreds of years" (Strong et al., 2009).

In effect, the CO_2 that in part causes climate change is removed from the atmosphere and sequestered in the ocean. Thus oceanographer John Martin quipped during a 1988 lecture: "Give me a half a tanker of iron, and I'll give you an ice age" (Martin, 1990).

After the early 1990s verification of the hypothesis that iron fertilization could indeed induce algae blooms, there was a sudden surge of media and commercial interest in the possibility of OIF to "solve" the climate crisis. A number of corporations were formed to invest in the process. These commercial ventures were in part driven by the goal of selling carbon offset credits—companies and individuals could buy "credits" that go towards avoiding or sequestering and equivalent amount of GHG emissions, in effect converting emissions into an economic commodity (Gillenwater et al., 2007). Throughout the late 1990s and early 2000s, nearly half a dozen venture companies gathered funds for OIF, performed small-scale demonstrations, and made plans to fertilize patches of ocean covering thousands or tens of thousands of square miles.

However, soon concerned scientists and environmentalists began to point out potential problems and side effects of these plans (Strong et al., 2009). For one, while the increased amount of algae would increase the depletion of carbon (a desired effect), it would simultaneously increase the depletion of other nutrients in the surface water that support other organisms in the food web. The process by which the algae absorbs carbon also produces organic matter, the decomposition of which can deplete the amount of oxygen in the water. These side effects can spread even beyond the fertilized area, potentially impacting ecosystems across the oceans. Moreover, this enhanced nutrient cycling (that occurs at depth as the produced organic matter sinks) can have rebound effects leading to an *increase* in nitrous oxide and methane—two greenhouse gases that produce the greenhouse effect more strongly than CO_2! Thus scientists began to conclude that the yield of carbon sequestered may actually be significantly lower than original hypothesized, and that OIF could cause widespread disruption of the ecosystem, among other potentially unpredictable downstream effects. While the topic is still under some discussion, many have come to the decision that these side effects and unknowns mean that OIF "should not be considered further as a means of climate mitigation" (Strong et al., 2009).

This history demonstrates how even well-meaning people often fail to understand the complex causal relationships found in complex systems, and how this failure could potentially have significant deleterious impacts. In an attempt to solve the climate crisis, scientists and corporations almost adopted actions that could have had devastating side effects because they were unaware of the complex web of causality involved in interactions among and between the global climate system and the ocean's ecosystem. In this way, humans' difficulty in understanding causality and complex systems can have dire repercussions in attempts to achieve environmental sustainability through individual actions and especially public policy. Indeed, Sterman and Booth Sweeney argue that these poor systems thinking skills have led the public in general to adopt an unsustainable "wait and see" approach towards global climate change:

> "[P]eople's intuitive understanding of even the simplest dynamic systems is poor. As long as people's common sense tells them that stabilizing emissions is sufficient there can be little political will or public pressure for policies that could stabilize climate and prevent further warming. As long as people believe the delays in the response of the system are short, they will conclude it is best to 'wait and see' if warming will occur and how much more harmful it will be before taking action. Such heuristics often work well in everyday tasks with low dynamic complexity... The same decision making heuristics that serve us well in simple systems may lead to disaster in complex dynamic systems such as the climate" (Sterman and Booth Sweeney, 2002).

As a group, people's failure to understand complex systems any apply systems thinking (rather than reductionist or linear thinking) has kept them from implementing policies that can help mitigate climate change—a failure that is in fact causing further harm through our inaction, despite our collective desire to help.

Indeed, this failure to enact effective policies for change results not only from a failure to understand causality and complex systems, but also from a failure to even realize that our understanding is flawed. We adopt well-meaning changes, unaware that such changes are either ineffective or even harmful to our goals. For example, Jevons Paradox (see Alcott, 2005) and the related the Khazzoom-Brookes Postulate (Saunders, 1992) proposes that increasing energy efficiency—a common action taken to increase sustainability—may actually lead to increased energy consumption as the cost of usage is reduced (see also Tomlinson et al., 2011). These rebound effects (Herring, 2008) result from indirect causal relationships of which people are unaware, and yet because of this unawareness people adopt potentially harmful changes anyway in their desire to support the public good. When it comes to complex causal systems such as environmental sustainability, humans also often demonstrate poor metacognition (Flavell, 1979)—poor knowledge of what they know or don't know. And because of this poor metacognition, people are unable to take effective actions or institute effective policies to achieve sustainability without causing further environmental impact through indirect, unexpected rebound effects.

Overall, humans often have difficulty understanding the multitude of cause-and-effect relationships in complex systems such as those surrounding topics in environmental sustainability, and this lack of understanding stymies attempts to enact broad changes while minimizing potentially harmful rebound effects. Yet if we want to improve on our understanding, we first need to perform the metacognitive step of identifying what we don't know—where our intuition and understanding of complex systems breaks down. Thus to achieve environmental sustainability we need to measure our understanding of the causal relationships in this complex system, to identify areas for further research and education to better enable effective policies and actions—getting people to not settle for a flawed "wait and see" approach.

1.1 Assessing Causal Understanding

In order to effect behaviors and public policies that can best achieve environmental sustainability, we need to assess people's understanding of the interactions within relevant complex systems. In particular, I am interested in assessing people's understanding of the *causal relationships* that make up these systems. Rather than focusing on a generic measurement of people's systems thinking skills, or their intuition about stocks, flows, and delays (à la Booth Sweeney and Sterman, 2000; Sterman and Booth Sweeney, 2002), I am considering how people identify and understand the ways in which different system components are causally connected. Such understanding directly relates to people's awareness of rebound effects that may occur—the more accurately people understand the interconnectedness of complex systems, the more likely they are to perceive unintended side effects that may be spatially or temporally displaced from their causes. Assessing causal understanding thus offers a foundation for identifying flaws in our knowledge or intuition that can alter the effectiveness of sustainability interventions.

There are some existing methods for assessing causal understanding, mostly coming from the domain of systems thinking education. Systems thinking pedagogies are assessed in a number of ways, including traditional qualitative and quantitative tests (e.g., Grotzer and Sudbury, 2000; Raia, 2005), interviews (Penner, 2000), and diagramming exercises (e.g., Booth Sweeney and Sterman, 2000; Gobert and Clement, 1999; Plate, 2010). Indeed, in a study comparing 5th-grade students' understanding of a complex system (specifically, plate tectonics), Gobert and Clement (1999) found that students who generated diagrams of the system outperformed students who produced written summaries of the system on post-test measurements of causal knowledge. This finding suggests that either diagrams may better match students' mental models of causal relationships in complex systems (enabling students to achieve higher scores), or that the process of diagramming may in itself help improve understanding. In either case, this and other previous work suggest that diagramming may be an effective method for assessing understanding of causal relationships.

Along these lines, most methods for considering causal understanding involve the construction of *causal maps* (Burgess et al., 1992; Chaib-draa, 2002; Fahey and Narayanan, 1989; Gopnik et al., 2004; Markóczy and Goldberg, 1995; Plate, 2006). A causal map is a type of concept map (Novak and Cañas, 2008)—a diagram spatially laying out concepts (often as boxes or circles) and then linking these concepts by the relationship between them (often as lines or arrows). In a causal map, the relationships drawn between concepts are cause-and-effect: which concepts are the causes of others (see Figure 1.1). Causal maps are usually drawn as directed graphs, where vertices are concepts and edges are the relationship between them. Edges are often marked as representing either an amplifying or dampening relationship—whether the causing concept leads to an increase or a decrease in the affected concept—often with a plus or minus sign respectively. Edges may also be given numerical weights indicating the relative strength of the relationship. These maps thus diagram the multiple causal relationships between concepts in a system; the creation of these maps is a form of cognitive mapping (Eden, 1992; Tolman, 1948). Analyzing cognitive maps has been positioned as an effective method for assessing and even enhancing causal understanding; as Boland and Tenkasi suggest: "Creating cognitive maps can reveal personal cause-and-effect logic, which in turn forces the individual to confront the reasonableness and validity of previously tacit cause-effect assumptions" (Boland and Tenkasi, 1995). Thus most existing methods for assessing causal understanding of complex systems involve the creation and analysis of causal maps.

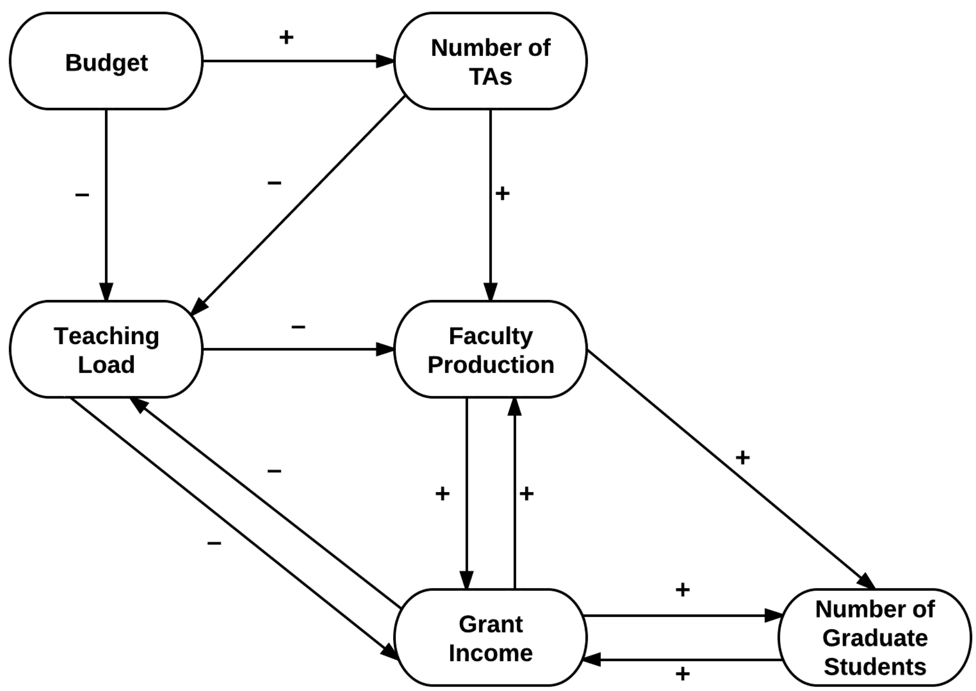


Figure 1.1: A causal map, describing the causes and effects of and on faculty production. Map adapted from Burgess et al. (1992).

One example of this method of assessment is work by Richard Plate (2006; 2010), upon which portions of this dissertation research are based. Plate used a framework called CMAST (Cognitive Mapping Assessment of Systems Thinking) to assess systems-oriented instruction of middle school and undergraduate students; a major component of this framework measures causal understanding, which is the focus of this dissertation. In Plate's studies, participants were asked to create a causal map based on their understanding of a short article describing "a fictitious fishing controversy based largely on the U.S. menhaden fishery" (Plate, 2010). In this article, the "samaki"[1] fish has great importance in the ocean ecosystem, and yet there is conflicting evidence as to whether the species may be in danger of being overfished. Students read the article, and then chose a number of concepts (written out individually on physical cards) out of a pre-made stack in order to build

[1]A fictional alias for the menhaden; see (Plate, 2006) and below for details.

their causal maps. The students were asked to identify the causal relationship between each pair of cards—either increasing (an increase in the first card causes an increase in the second, or an amplifying effect), decreasing (an increase in the first card causes a decrease in the second, or a dampening effect), or neither (if the concepts were not causally related). Participants then glued the cards down onto a large piece of paper, hand-drawing the directed edges of the causal graph between the cards and labeling each edge with a "plus" or "minus" symbol to indicate direction. The relative weights of the relationships were not considered.

Plate then used these maps to assess each student's causal understanding. He measured student accuracy by comparing the causal map to expert-generated maps on the same topic, scoring each relationship with a rubric based on how many experts agreed that the relationship existed (see Plate, 2010, for details). For example, relationships identified by 4/4 expert groups received 8 points, while relationships not identified by any expert earned -0.5 points to discourage guessing. In this way, students were given a total score based on how accurate and complete their maps were in comparison to expert knowledge. Plate also measured link density (how connected the concepts were to one another) via the graph theoretic degree of each vertex in the causal graph, in order to assess the linearity of the generated map. This research found that systems thinking instruction increases both the accuracy and connectedness of the causal maps, and demonstrated how this form of cognitive mapping process can be used to assess the accuracy and complexity of causal understanding.

However, while this cognitive mapping method is able to measure the accuracy of a person's causal understanding of a system, it suffers from a number of limitations. For one, such methods require significant time and effort to perform; causal maps are often manually constructed by expert researchers (after Axelrod, 1976), or otherwise involve coding and scoring by hand. While CMAST can be used to assess an entire classroom in about an hour, scoring the resulting maps requires days to analyze the entire collection (Plate, personal communication, July 16, 2012). For this reason, such assessments can only be performed on a relatively small number of subjects—these techniques cannot feasibly be used to assess causal understanding of very large groups of stakeholders.

Moreover, these assessments focus only on the final result of the causal mapping exercise, rather than engaging with the *process* of constructing a causal map. Diagramming causal relationships may support greater measurements of causal understanding (Gobert and Clement, 1999) because subjects improve their mental model of the causal system (which is then reified in the diagram) through the construction process—by explicitly considering how components of a system are causally related, subjects may discover and acknowledge heretofore hidden causal relationships. In addition, because causal maps are generally reified in a *de facto* linear sequence, considering this sequence may support a more robust analysis of how the subject understands the causal system. For example, analysis of the sequence or process could suggest which relationships are perceived more prominently (and therefore were built first), or may act as catalysts for the map's construction by leading to the inclusion of other relationships. Nevertheless, the logistical difficulties of manually tracking and assessing the timing and sequence of causal map construction restrict the use of these potentially informative data as a component of causal assessment.

Interactive computer systems may be able to help overcome these limitations. At the simplest level an interactive computer system can mirror the CMAST technique, having subjects generate causal maps on their computer screen, rather than on paper. While this technique may lose some of the benefits of tangible interaction (see e.g., Marshall, 2007), it enables generated maps to be scored automatically rather than needing to be processed by hand[2]. Indeed, such automatic scoring has become a staple of assessment methodologies, particularly in standardized tests ranging from grades K-12 to the GREs required for admission to many graduate schools. In addition, interactive computer systems could more readily gather data about the process through which subjects construct causal maps, recording interaction events that occur within the system (e.g., which relationship is created when)—data that can then be integrated into the automated assessment. In these ways, computational systems may be able to reduce the time required to evaluate produced causal maps,

[2]Maps can also be assessed using a combination of computerized scoring and manually verification, if desired.

while effectively collecting further data on the process of map construction for use in assessment.

Moreover, drawing from the fields of serious games (e.g., Abt, 1970; Sawyer and Smith, 2008) and gamification (e.g., Deterding et al., 2011a), I am particularly interested in how video games as a form of interactive computer system may be able to further support assessing causal understanding. Games are themselves highly complex dynamic systems (Salen and Zimmerman, 2003) in which a large number of variables and rules interact to produce emergent behavior, and with which players eagerly and voluntarily engage. Proponents of games for education argue that good games provide their own intrinsic motivation (games are fun), and offer a safe environment for players to experiment with and learn to understand the complex system that is the game. Indeed, Gee (2003a) suggests that learning to play a game involves learning to master the behavior of the system—and it is the process of gaining this mastery that makes playing the game fun (Koster, 2005). Proponents argue that this link between games, fun, and learning positions games as highly suitable for education, and as such games have a long history of use in educational contexts[3] (see Hung, 2011).

Yet despite this history, there has been less consideration of how games may be used for assessment (though see e.g., Shute et al., 2011). Assessment is a core component of most educational interventions, and as such it is possible that the benefits of games for education (e.g., high engagement, allowance for experimentation, adaptive challenges—see Gee, 2003a) might also apply to assessment. For example, using games for assessment could potentially increase engagement with the evaluation process and encourage subjects to "try harder" and be more thoughtful or attentive in completing the assessment, allowing the assessment to better measure understanding. Similarly, gamifying assessment and making it "fun" could potentially reduce test anxiety (Hembree, 1988), thereby supporting accurate assessment of a wider variety of subjects.

Furthermore, previously developed games have been able to successfully elicit human knowledge

[3]Including systems education: the MIT Beer Game (described in Chapter 3) has been used in testing and teaching systems thinking since the 1960s, and has been deployed as an online computer game (Jacobs, 2000; Nienhaus et al., 2006).

and understanding in much the same way that cognitive mapping assessments do. These *games with a purpose* (GWAPs; see von Ahn and Dabbish, 2008) ask players to apply their understanding of virtual models both simple (e.g., the content of a picture in the *ESP Game* (von Ahn and Dabbish, 2004)[4]) and complex (e.g., the structural balance of protein folding in *Foldit*). A similar approach may be used to solicit understanding of complex causal systems; instead of identifying images, players could identify causal relationships. In this way, existing successful systems suggest games as a particular form of interactive system may effectively support assessment with a variety of benefits in terms of both data collection and engagement.

1.2 Summary and Research Contributions

This dissertation research explores how interactive computer systems can support assessing understanding of complex systems through consideration of the process of causal mapping. Furthermore, it considers how computer games as a particular form of interactive system may support assessment. These two concerns provide the basis for this work's two main research questions:

1. What can interactive systems reveal about the structure of causal maps and the process by which people construct these maps as reifications of their understanding of complex causal systems?
2. What characteristics of games as a form of interactive systems support or impede assessment, particularly of causal understanding?

To answer these questions, this dissertation presents a prototype interactive computer system in the form of a game called *Causlings*, which works to assess causal understanding. This system is based on Plate's CMAST framework (Plate, 2006, 2010) for assessing causal understanding through cognitive causal mapping: I apply his assessment techniques through the medium of an

[4]The ESP game was later turned in "The Google Image Labeler", and was used to tag a significant percentage of the images indexed by the search giant.

interactive computer game. Though this prototype, I analyze automatically gathered data about the process of causal mapping for insight into players' systems thinking, while also considering the effects of the game-like framing and the use game dynamics on the causal assessment process. I further explore these questions through the results of a user study testing this prototype with both students and non-student subjects.

With this system and study, this dissertation offers the following key contributions:

1. Procedurally evaluating the structural forms of causal maps created in an interactive system might suggest the presence of reductionist thinking.
2. The sequence in which a user constructs a causal map can be analyzed to detect the presence of linear thinking, and this analysis can be automated by an interactive computer system.
3. Although games as interactive computer systems can be effective in enabling learning, they may be less readily effective in supporting stand-alone assessments due to (a) requiring an *a priori* understanding of the complex game system used in the assessment, and (b) traditional educational assessment contexts not supporting the forms of feedback critical to supporting game-based learning.

This work has implications for how educators teach and assess understanding of complex causal systems–suggesting a move away from the prominence of linear narratives both in curricula and in the development of future games for education and assessment.

1.3 Outline of this Research

The remainder of this dissertation proceeds as follows:

Chapter 2. Background and Related Work: Causality surveys previous research in the domains of causality, systems thinking, and cognitive mapping. It describes a theoretical framework and set of semantics to use in discussing the assessment of causal relationships.

Chapter 3. Background and Related Work: Games continues my survey of previous research. I discuss related work in the fields of games for education and gamification in general, considering potential benefits and limitations of applying game dynamics to assessment.

Chapter 4. Causlings *Implementation* details the design and implementation of the *Causlings* game prototype. It explains how the interactive system functions as an assessment tool and can automatically collect data on the map building process, as well as offering justifications for game design decisions.

Chapter 5. Evaluating Causlings describes the study procedures used to explore the data collected by the *Causlings* game prototype and used to consider the impact of the game form on assessment. This chapter presents hypotheses and variables of interest.

Chapter 6. Results and Analysis reports the data collected from the user study, along with an analysis of these data.

Chapter 7. Discussion reflects on the results of the user study and the design of the prototype system as a whole. It addresses potential implications for those results, explaining their significance in terms of assessing systems thinking and designing educational games dealing with complex causal systems. This chapter explains and justifies the dissertation's contributions.

Chapter 8. Future Work lists potential extensions to this research, including study limitations to be addressed.

Chapter 9. Conclusion Summarizes the key findings and contributions of this dissertation.

Chapter 2. Background and Related Work: Causality

In this and the following chapter, I detail the theoretical background behind this dissertation's study of systems for assessing causal understanding. This chapter explains "causality" for the purposes of this work, discusses the relationship between causality and systems thinking, and provides a set of semantics for discussing causal relationships. It also describes previous work in causal mapping that lays the groundwork for the development of this thesis.

The research presented in this dissertation touches on a variety of well-trodden research domains, including the philosophy of causation, systems thinking, complexity theory, Bayesian networks, cognitive mapping, and pedagogical assessment, and theories of games and play. Each one of these could and has been the subject of multiple dissertations in their own right. As such, I do not attempt to cover each of these fields in depth; rather I aim to provide brief summaries of these domains and point to relevant concepts that inform my research.

2.1 What is Causality?

Causality (also synonymously called causation) is an attempt to answer the "why" of a natural phenomenon. Why does a light bulb illuminate when plugged into a battery? Be*cause* of the current flowing through the wire. Why are populations of Sumatran tigers shrinking? Be*cause* of habitat loss, which in turn is caused by over-logging and increased palm oil production (see e.g., Fitzherbert et al., 2008; Linkie et al., 2003). The relationship between *cause and effect* is a way that people can assign meaning to phenomena or events—a cause is a phenomenon that brings about an effect (another phenomenon). Causality allows people to understand events in relationship to one another, to make sense of and act within the world around them:

> "Causal knowledge is important for several reasons. Knowing about causal structure permits us to make wide-ranging predictions about future events. Event more important, knowing about causal structure allows us to intervene in the world to bring about new events—often events that are far removed from the interventions themselves" (Gopnik et al., 2004).

As humans, our understanding of causality thus provides a mental model for how natural phenomena interact, and how our actions will affect our environment (i.e., what are the effects we will cause).

2.1.1 The Metaphysics of Causality

Philosophers have argued about the theory and definition of causation for millennia (see e.g., Hulswit, 2004; Schaffer, 2008, for summary and examples). Causality is a central topic in metaphysics, and thus there is a long philosophical tradition of investigating what it means for something to be a cause of something else[1]. For example, Aristotle identified four different types of causes or ways of explaining an object or phenomenon: the *material cause* (the raw materials, e.g., the bronze of a statue), the *formal cause* (the plan or mold, e.g., the shape of a statue), the *efficient cause* (the external entity that starts the changing, e.g., the artisan), and the *final cause* (the sake for which a thing exists, e.g., a statue may be cast to honor the subject) (Falcon, 2011). Even this ancient view posited a variety of ways of thinking about causality, including a teleological one in which the purpose of the phenomenon (in a way, the effect) is the cause of that phenomenon. Aristotle's

[1] In order to better scope the discussion vis-à-vis its intended audience, my survey focuses primarily on causality as considered in Western philosophy. No slight is intended through the omission of non-Western thought and traditions.

theory suggests that there are multiple components (the bronze, the mold, the artisan, the purpose) that all interact to bring about an effect; indeed, his defense of the final cause suggests that only by looking at the result can we understand the process—an argument with analogs to the systems idea of emergence (see below).

On the other hand, modern Western philosophy has questioned the basis on which we perceive causality—how we know for sure that phenomenon A caused phenomenon B. For example, Hume argued in *An Enquiry Concerning Human Understanding* (1748) that we have no sensory capabilities to detect causality (in the way that we have sensors for sight and smell), and thus only identify causality by developing habits of mind. We come to assume that if two phenomena are contiguous in time and space, and one always occurs before the other, then they are causally linked such that the first phenomenon always causes the second—even though we have no basis for claiming that this will always be the case. Indeed, the question of *necessity* (whether given A, B must necessarily occur as an effect) is one of the fundamental arguments in the history of causation (see Hulswit, 2004). Attempts to answer the question of how we *know* for sure that A caused B in part drove the development of current scientific and statistical practices; in scientific research we usually treat A as causing B if, when all other variables are constant, a change in A leads to a change in B (Woodward, 2003). (Such a claim can also be defined probabilistically, where A is a cause of B if it increases the probability that B occurs—see Hitchcock (2007) for an example definition). Cause and effect are thus temporally ordered, though theoretically the temporal distance between cause and effect may be quite large.

However, identifying necessity in causation is further complicated when we consider the interrelationships of multiple variables, as is the case in most natural systems. We may attempt to explain causal relationships counterfactually (e.g., if A had not occurred B would not have occurred, so A caused B), but run into problems with the possibility of preempting causes. Consider the following riddle:

> A man sets out on a trip across the desert. He has two enemies. One of them poisons the traveler's water. The other (not knowing this) makes a hole in the bottom of the traveler's canteen. The poisoned water leaks out and the traveler dies of thirst. *Who caused the traveler's death?* (Adapted from Mackie, 1980).

Holding all other variables constant, both enemies increased the probability of the traveler dying, and so both caused the traveler's death. However, because the actions of one enemy preempted the other, we may not wish to blame them both. Or should the traveler be responsible for his own death, as he made his enemies wish to kill him? Thus the answer to the riddle is open to philosophical debate (see also Menzies, 2009); even this simple example demonstrates some of the difficulty in formalizing general theories of causation. Indeed, arguments over such riddles are representative of people's potential to misunderstand or hold conflicting interpretations of the causal "source" of a particular effect—particularly in multivariate complex systems.

Yet despite the complications of developing formal, metaphysical definitions of causality, people are able to understand causal relationships between phenomena from a very young age. While developmental psychologist Jean Piaget believed that young children had "precausal" understanding (Piaget, 1930), a wide variety of recent studies have shown that even infants are able to perceive and expect physical causality, such as when billiard balls collide and launch one another (see the introduction to Gopnik and Schulz, 2007, for one review). Nevertheless, psychologists have not reached an agreement on how exactly infants learn to make these causal connections, or how an understanding of physical causality can be extended to an understanding of causality in more abstract scenarios or systems. Along similar lines, computer scientists from the fields of artificial intelligence and robotics (e.g., Pearl, 1988) are still working on developing computer systems that can perform causal reasoning in ways that seem innate to human infants. In short, formally defining and understanding causality and causal reasoning is an unsolved problem in a wide variety of disciplines. This difficulty in modeling causal understanding informs the difficulty of assessing causal understanding, which often occurs through explicit modeling.

Nevertheless, the difficulties of formally modeling causality do not usually restrict people in their everyday actions and understanding. As John Sowa points out:

> “[P]eople, animals, and even plants benefit from causal predictions about the universe, independent of any conscious thoughts they might have about causality. People plan their daily trips to work, school, or shopping; squirrels bury nuts for the winter; and plants turn their leaves upward and their roots downward. The success of these activities does not depend on a theory of physics that is accurate to seven decimal places, but it does imply that the universe is sufficiently regular that even a rough rule of thumb can be a useful guide for one’s daily life” (Sowa, 2000).

People rarely require a strict formal understanding of causation, instead relying on a common sense understanding of the often-thorny relationships between causes and their effects that is "valid at the level of human experience" (Born, 1949). With this in mind, for the purposes of this research I will be considering causality in its more general, every-day sense: causality is "the relationship between a set of factors (causes) and a phenomenon (the effect). Anything that affects an effect is a factor of that effect"[2]. For this dissertation, causation occurs whenever one factor influences another, regardless of circumstances. Yet even this common-sense definition contains certain assumptions (e.g., about the inclusion of indirect effects or the lack of limits on temporal distance)—and it is these unrealized assumptions that may lead people to misunderstand the causal influences of complex systems.

Note that a full deconstruction of the philosophical and psychological meanings of causality is beyond the scope of this dissertation. In this research, I will not focus on the metaphysical issues surrounding causation, beyond their implications for the difficulties in modeling and fully understanding complex causal relationships. Yet at the same time, I acknowledge that this understanding rests on philosophical assumptions and beliefs that have significantly influenced Western culture; indeed, the difficulties of accurately modeling causality may be based on aspects of cultural philosophy which subjects have (perhaps unconsciously) internalized.

[2]This definition is adapted from Wikipedia, representing an aggregate of general knowledge and understanding.

2.2 Causality and Systems Thinking

In this dissertation, my consideration of causality is based on the concept's usage in *systems thinking*. In systems thinking, the locus of interest is not just an individual element (e.g., a person, a biological cell, an electron), but a whole *system* of potentially heterogeneous elements. Ludwig von Bertalanffy, one of the founders of systems theory (the study of systems in general), defines a system as "a complex of interacting elements. Interaction means that elements, *p*, stand in relations, *R*, so that the behavior of an element *p* in *R* is different from its behavior in another relation, *R'*" (von Bertalanffy, 1969)[3]. Or less formally, a system is "a group of elements standing in interrelation with themselves and with the environment" (von Bertalanffy, 1969). The crux of this definition is how elements *interrelate*—how elements interact with the other elements in the system, and how those interactions are influenced by other relationships within the system. Gregory Bateson (1972) describes a systems view as considering verbs (interactions) rather than nouns (elements).

Indeed, causality in systems theory can be defined as the interactions that involve some set of elements (the inputs or causes) affecting the future state of other elements (the outputs or effects)—see (Mesarović and Takahara, 1975; Windeknecht, 1967) for formal mathematical definitions. Broadly, from a systems perspective, causation occurs when an element interacts with another element in order to change it in some way; if element A affects another element B, then A has caused a change in B, and thus there is a causal relationship between A and B. In this sense, the phrase "A causes B" can be read to mean: "A causes a change in B"; the exact nature of the change need not be specified (but see below). Moreover, because these interactions all influence one another simultaneously, the relationships between elements can give the system as a whole behaviors or properties that are not found in the individual components. This phenomenon of unexpected behavior is called *emergence*, and is a fundamental property of *complexity*.

[3]A system can also be defined mathematically, such as through a set of differential equations. Indeed, the goal of von Bertalanffy's general systems theory is to identify mathematical principles that can be applied to systems across domains (von Bertalanffy, 1972).

2.2.1 Emergence and Complexity

Consider the behavior of a flock of birds flying through the air. The birds are able to head in the same direction and perform graceful group turns without mid-air collisions. One hypothesis may be that careful communication or a form of hierarchical leadership enables this behavior; however, the use of these mechanisms is not supported through observation (Potts, 1984). On the other hand, computer simulations (Reynolds, 1987) have been able to model this complex and beautiful behavior using just three simple rules:

1. Collision Avoidance: avoid collisions with nearby flockmates.
2. Velocity Matching: attempt to match velocity with nearby flockmates.
3. Flock Centering: attempt to stay close to nearby flockmates.

By having each individual bird follow these rules (giving earlier rules precedence), the group as a whole demonstrates the flocking behavior. Flocking is thus an *emergent behavior* (Holland, 1999)—a complex behavior that arises from a set of simple rules; one that is greater than the sum of its parts[4]. From a systems perspective, each element of the system (a bird) uses these rules to interact with the other elements, and thereby produces a new behavior displayed by the system as a whole. Indeed, these rules for element interaction can also model the behavior of schooling fish or herding mammals, allowing this idea to generalize to other domains. Discovering how these individual interactions influence and relate to the behavior of the whole system is the overall goal of systems thinking.

Such emergence is a fundamental component of *complex systems* (e.g., Waldrop, 1993), which are the systems of interest in this dissertation. Indeed, complex systems are defined as those in which a person cannot predict their emergent properties simply by considering their individual elements. While a full treatment of complexity theory is outside the scope of this research, it is worth noting that these kinds of emergent systems display what Senge (1992) calls "dynamic complexity"—

[4]Conway's *Game of Life* (1970) provides another classic example of emergent behavior.

complexity that arises from the interactions between elements, rather than from the sheer number of elements. It is not the scale of systems that makes them complex (though scale is a distinct factor in the difficulty of intuiting about systems such as environmental sustainability; see e.g., Tomlinson, 2010), but the unexpected behaviors and causal relationships that are embedded in the system.

Because in complex systems emergent behaviors arise from the interactions of multiple, potentially unrelated elements, a change in a single element can *indirectly* lead to an emergent change in another. A causal relationship between two elements may not necessarily correspond to an interaction between those elements–instead, the causal relationship may represent an indirect action leading to an emergent behavior. For example, a bird flying on the left side of a large flock may cause a bird flying on the right side to change its heading, even though those two birds did not directly interact. The interconnectedness of the system means that causal relationships do not occur in isolation (as they may in metaphysical scenarios or well-designed psychological experiments), but within a complex system of other concurrent causal relationships. Emergence in complex systems means that an effect can be brought about by the interaction of apparently non-related causes—and in fact, all emergent behaviors can be considered as being caused by indirect interactions of some kind.

The possible disconnect between element interaction and element causation can make it difficult for people to detect and intuit about causality in systems. Recall from the beginning of the chapter: "knowing about causal structure allows us to intervene in the world to bring about new events" (Gopnik et al., 2004). Yet in complex systems with emergent properties, the effects of a cause need not explicitly follow from the interactions between elements but can instead emerge from the sum of these interactions. Interacting with one element may not cause an expected change in that element, by may instead cause some other emergent behavior.

Furthermore, dynamic complexity means that even systems that seem simple (in that they have only a few elements) can actually demonstrate quite complex emergent behavior. One classic example of such dynamic complexity is the Lotka-Volterra predator-prey model (see Berryman, 1992), which has only two elements: the population of predators (e.g., wolves) and the population of prey (e.g., sheep). In this model, the growth of the wolf population is dependent on the population of sheep, and vice versa. As wolves eat the sheep and increase their population, the sheep population drops because there are more wolves eating. But as the sheep population drops, there is less food for the wolves, and so their population also begins to drop. And with fewer wolves to eat the sheep, the sheep population can begin to rise again. This means that there are more sheep to feed a growing population of wolves, and the cycle continues (see Figure 2.1). Thus even simple relationships between the elements (the growth of the predator and prey populations are inversely proportional) can produce dramatic and complex behavior due to dynamic complexity. Systems with even more elements or with more intricate relationships between elements (as in environmental sustainability) produce emergent behaviors that are even more difficult to predict.

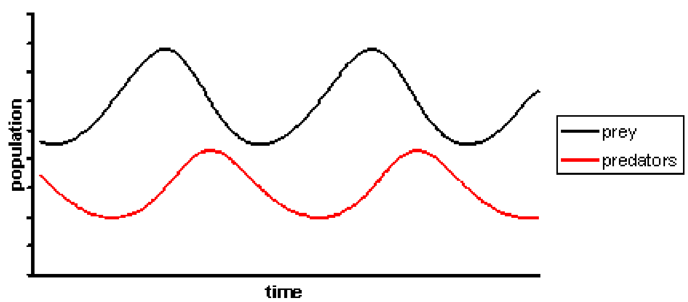


Figure 2.1: An example cycle of the Lotka-Volterra predator-prey model. Graphic used with permission from Wikimedia Commons.

This form of dynamic complexity demonstrates another property of the causal relationships of complex systems: their non-linearity. Systems display *non-linear causality*, a concept at the heart of systems thinking. Rather than having a system in which each cause has a single effect and each effect has a single cause, the emergence of complex systems ensures that any single cause (e.g., interaction with an element) will likely have multiple effects, and that certain effects (e.g., emergent behaviors) are caused by interactions among multiple components. The change in the predator population does not have a linear relationship to the change in prey population–instead the two elements have a mutual, non-linear cause-and-effect relationship. In this way, the emergent behaviors of complex systems cannot be modeled through simple, linear causal relationships, but rather need to be considered as an entire web of non-linear interactions. (Note that systems can also have non-linear causal relationship in the sense that the effect is not linearly proportional to the cause. A small change in element A can cause a large change in element B, and vice versa). As Plate says, non-linear causality is the "shifting from the fundamental model of a causal chain to that of the causal web... one cause can have more than one effect, and one effect can be the result of more than one cause" (Plate, 2006). Systems thinking is based on this idea of non-linear causal relationships, such as exist in emergent, complex systems.

This predator-prey model is also an example of a self-regulating system—a system that balances itself through a feedback loop. Self-regulating systems and the feedback loops that enable them are commonly the focus of systems thinking, as well as forming the basis for fields such as cybernetics (Wiener, 1948; see also Hayles, 1999; Mayr, 1986 for history of the field). A *feedback loop* takes the ultimate effect of a chain of indirect causal relationships and turns it back into a cause at the beginning of the chain, thus creating a cycle. Such *causal loops* can be negative feedback loops (i.e., balancing loops)—such as the above predator-prey model, or how a thermostat causes the heat to turn on when below a target temperature but not when above. Causal loops can also be positive feedback loops (i.e., reinforcing loops)—such as how increases in global temperatures cause polar ice caps to melt, but this melted water absorbs more of the sun's energy than solid ice and so causes global temperatures to rise, which continues to cause the polar ice caps to melt (see Figure 2.2).

Causal loops are an example of non-linear causality, representing a kind of mutual interaction among system components. Indeed, causal loops are part of what enables the non-linearity of changes in complex systems (Richmond, 2001): a small cause can start a positive feedback loop and turn into a big effect, or a large cause can hit a negative feedback loop and not produce much of an effect at all. Causal loops represent a common form of causal relationship in complex systems, and contribute to making the causal webs of complex systems difficult to predict or intuitively

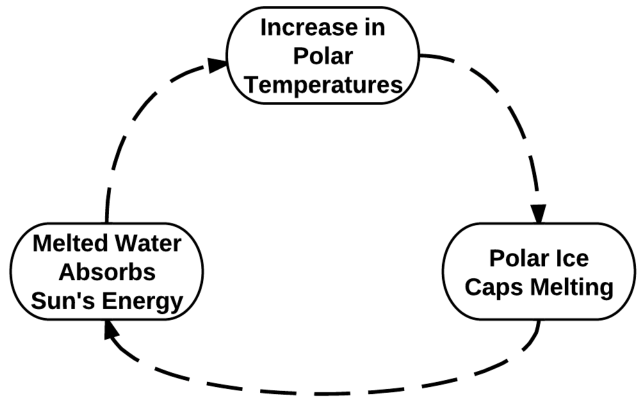


Figure 2.2: A positive feedback loop. This example is adapted from Plate (2006), who adapted it from former Vice President Al Gore.

understand.

In these ways, the non-linear interactions of components in complex systems leads to indirect effects and emergent behaviors; the goal of systems thinking is to understand these relationships and phenomena. Note that overall, a systems perspective of causality focuses more on the range and variety of often-unexpected effects that occur within complex systems than the metaphysical questions of causation. Systems thinking primarily involves considering how multiple causal relationships interact with one another. As such, when considering complex systems in this dissertation I consider an element to "cause" another if it effects a change in that element—in a way, using solely Aristotle's *efficient cause* (the external actor that initiates change). This thesis focuses on using systems thinking to identify the non-linear causal relationships that bring about such emergent behaviors.

2.2.2 Systems Thinking vs. Reductionist and Linear Thinking

The properties of emergence and non-linearity are fundamental to the consideration of the complex systems discussed in this dissertation. Systems thinking enables a person to properly understand these features–an understanding that may not be apparent from classical scientific approaches that rely on *reductionist thinking* or *linear thinking* (see Plate, 2006; von Bertalanffy, 1972, for more details). Reductionist thinking is an approach to understanding and problem solving that involves dividing a concept into smaller component parts that can be understood more readily, and then mentally recombining those components into an understanding of the whole. A reductionist approach can help people to deal with the potentially overwhelming complexity of large-scale systems. Reductionism has deep historical antecedents, and is often at the core of conventional educational curriculum (Plate, 2006). Much of modern science and science learning is based on reductionist approaches: biology students study cell components in order to understand cells, cells to understand organisms, and organisms to understand entire ecologies (in fact, the very concept of biological classification stems from a reductionist view). Young children are even taught to read using reductionist approaches: sounding out each letter and then combing them into words. Such an approach allows people to easily study intricate systems and phenomena by providing an almost algorithmic recipe for thinking: break the concept down until you reach a level you understand. Learning through divide and conquer.

However, the concept of emergence means that the emergent properties of complex systems are not susceptible to reductionist thinking—one cannot identify these properties by considering the individual parts. Instead, the system needs to be considered *holistically*, as through systems thinking. Taking each of the Reynolds' flocking rules (above) in isolation does not necessarily indicate that a flock of birds would be able to gracefully turn—all the rules have to be considered in tandem. Training people to understand emergent properties thus requires encouraging an increase in systems thinking and a decrease in reductionist thinking; for this reason, the two modes are often positioned as contrasting with one another. Indeed, as previous research (e.g., Moxnes, 2000) has indicated, part of the challenge with dealing with issues involving complex systems, such as environmental sustainability, is that considering these systems through a reductionist approach does not always work—individual components cannot necessarily be considered separately from the whole. As one example, considering the energy efficiency of a machine independent from how energy is consumed is what leads to Jevon's Paradox (Alcott, 2005), in which efficiency actually increases consumption. The very concept of complex systems such as sustainability implies that their parts are interconnected and must be considered holistically. One cannot only consider a small number of the pieces that influence a complex system in environmental sustainability. This is not to say that there is no use for reductionist thinking, or that the use of systems thinking precludes reductionist thinking (indeed, reductionist thinking may be much more effective when considering non-emergent behaviors)—rather that understanding the kinds of complex causal systems of interest in this dissertation relies on systems thinking, not reductionist thinking.

Similarly, the non-linearity (i.e., multiplicity) of the causal relationships in complex systems means that systems thinking is often more effective in understanding such systems than linear thinking. Linear thinking involves the consideration of a system or concept as a single step-by-step process, where each step is completed before the next is begun (in computer science terms: serial processing rather than parallel processing). Indeed, such linear thinking is also commonly found in basic science education, particularly in the narratives used to explain scientific process: for example, the water cycle is often implicitly framed as a step-by-step sequence of events. Linear thinking is related to reductionism in that it entails breaking a system down into component parts; however, the emphasis when describing thinking as linear is on the strong sequential process of problem solving or understanding. In a causal system, linear thinking organizes causal relationships into a series or chain: one component causes another effect, which causes another effect, which causes another effect, and so on. Although such thinking is able to capture indirect causal relationships (in the form of causal chains; see below), such step-by-step consideration is in stark contrast to the more holistic systems view that supports an understanding of multiple, non-linear causations (Richmond, 2001).

Linear thinking is quite common in considerations of causality; for example, people often describe their understanding of causality in terms of centralized linear narratives (Grotzer and Perkins, 2000; Plate, 2006), where the causal relationships are like the events in a story[5]. Indeed, some

[5]Indeed, English novelist E.M. Forster points out how engaging narrative plots (rather than simple stories) are intricately linked to causality: "A plot is also a narrative of events, the emphasis falling on causality. 'The king died and

believe that narrative is fundamental to how people make sense of the world around them, organizing their understanding of their lives and selves into stories (e.g., Bruner, 1991). In this view, being able to accurately frame a concept in a narrative can be used as a strong indicator for assessing understanding, although such narratives imply a linear form. Thus understanding complex, non-linear systems through systems thinking rather than linear thinking would require the need to construct *non-linear narratives* of the causal relationships: focusing not on the "chains" of causal relationships but on the "web" of multiple interconnected causalities.

Reductionist and linear thinking are both effective problem solving strategies, and may be instinctive or intuitive approaches to understanding (though such intuitiveness may be a side effect of the educational and cultural infrastructure that trains and emphasizes these skills). Nevertheless, the emergent, non-linear properties of complex systems that make them complex is resistant to classical, reductionist science. Instead, systems thinking approaches are necessary for understanding the complex causal interactions found within these systems. As such, this dissertation considers how such forms of systems thinking may be assessed, in order to measure understanding of complex causal systems.

2.3 Causal Semantics

This dissertation presents an interactive computer system for assessing people's understanding of causal relationships in complex systems. To better explain this system and its mode of assessment, in this section I present a set of semantics for discussing causal relationships, including the parameters by which I will be measuring a particular relationship. These parameters are synthesized from previous work in classifying causality (primarily Hitchcock, 2007; Kadaba et al., 2007; Schaffer, 2008) to suit the purposes of this research. Following Hitchcock's (2007) example, I begin by de-

then the queen died,' is a story. 'The king died, and then the queen died of grief,' is a plot. The time sequence is preserved, but the sense of causality overshadows it. Consider the death of the queen. If it is in a story we say 'and then?' If it is in a plot we ask 'why?'" (Forster, 1927).

scribing an analogous example of a small (not necessarily complex) system of causal relationships, using this example to help illustrate the parameters of interest.

Consider the effects of public transportation on fossil fuel usage (and its greenhouse gas emissions). A common suggestion (e.g., EPA, 2011) for people to reduce their everyday fossil fuel usage is to use public transportation such as local buses when they need to get somewhere. Driving a car is normally a significant source of fossil fuel usage, but taking public transportation causes cars to be driven less (because the drivers are instead on the bus). Thus taking public transportation reduces driving, and thereby reduces the fossil fuel usage caused by that driving. But at the same time, buses and other forms of public transportation often directly use fossil fuels themselves—though presumably not as much as would otherwise be used by the cars that the buses replace. While public transportation may directly use a small amount of fossil fuels, using such transit should lead to an overall reduction in fossil fuel usage and the associated greenhouse gas emissions. This system and proposition is represented visually in Figure 2.3, using a causal map (see below).

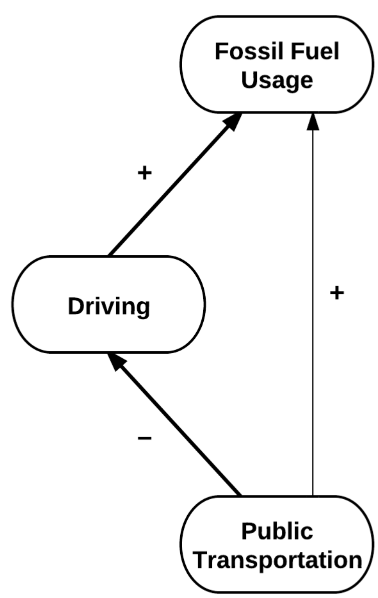


Figure 2.3: Effects of Public Transportation on Fossil Fuel Usage.

2.3.1 Relata: Factors

The first component of this system to consider is the elements within the system—in this case: "public transportation", "driving", and "fossil fuel usage". These elements are Bateson's (1972) nouns in the system, and they act as the relata for the causal relationships that make up the system.

An every-day view of causality often considers relationships in terms of events (e.g., the event of throwing a rock causes the event of a window breaking). But because the example system reflects a general set of relationships, it does not talk about the specific event of "a person takes public transportation", but rather "public transportation" considered as a more general, abstract concept. Relata need not be events or even discrete objects (such as a particular rock or billiard ball), but can instead be states of affairs, properties, aspects, features, descriptions, facts, or other forms[6]. Indeed, the relata of the systems discussed this dissertation can include any mix of these forms. As such, I will refer to them as the *factors* of a causal relationship, allowing the discussion to abstractly consider a heterogeneous set of relationships within a complex system, rather than restricting the modeling of complex systems to a particular kind of noun. This abstraction process may better enable the assessment of potentially divergent understandings of a system. When displayed graphically, factors may also be referred to as *nodes* in a causal map.

Note that one can often use natural language to try and clarify the factors of a causal relationship. For example, one might clarify "fossil fuel usage" by relabeling it as "fossil fuels burned" (as opposed to, say, "fossil fuels mined"). Such relabeling can potentially affect the meaning of the relationships and the reader's understanding of them. Because I consider factors in a more generalized sense in order to include a wider variety of causes in modeling a complex system, the meaning and specification of these factors can rely on human language usage. While the meaning of such language is constantly being constructed and is open to misinterpretation (Goffman, 1959), my consideration of causality at the "level of human experience" (Born, 1949) makes this level of

[6]Schaffer (2008) describes how there is support for considering causation in terms of any of these forms.

specificity appropriate for this research.

Furthermore, specifications of causality often rely on the assumption that the cause factor occurs before the effect factor. In complex systems using more abstract factors, this timing can be vague and harder to determine—in particular, complex systems often involve continuous interactions rather than discrete events. The concept of delays in dynamic systems (see Booth Sweeney and Sterman, 2000) means that the existence of a causing factor (such as public transportation) may overlap with or be significantly removed from its effect factor (such as reduced fossil fuel usage). It is often difficult to view factors in a causal relationship as events occurring at distinct times[7], particularly when such factors refer to generalized or communal behaviors as common to discussions of sustainability.

2.3.2 Relationships

As explained above, I consider two factors to be causally connected (that is, there is a causal relationship between them) if a change in one factor leads to a change in some aspect of the other. A person taking public transportation for a particular trip is not driving; therefore a change in public transportation usage can be said to cause a change in driving and there is a causal relationship between these factors.[8] Note that the stability of a relationship may be in question: if driving only sometimes changes fossil fuel usage (e.g., because of the advent of solar-powered vehicles), then the necessity requirement for causality—a change in A always causes a change in B—would be voided. In this research, I do not require total stability because of the complexity of the systems under consideration; thus a conditional is implied in causal relationships: driving *can* cause fossil fuel usage.

[7]Questions of time travel and causation are *definitely* beyond the scope of this research.

[8]There may of course be mutual causality between these factors—a change in driving habits can influence public transportation usage, and vice versa. This simple model assumes that driving is the "default" mode of transportation (as is the case in much of Southern California), and thus public transportation can be viewed as the dependent variable and the source of the causation.

Causal relationships have a number of parameters of note reflecting the nature of the change in factors, detailed below.

2.3.2.1 Causal Direction

Are we talking about how public transportation affects driving? Or how driving affects public transportation? This is a question of the *direction* of the causal relationship—the distinction between which factor is the "cause" factor and which is the "effect" factor. Yet there is also another form of directionality that must be considered in a causal relationship: whether the cause factor promotes or prevents the effect factor. Does public transportation promote, encourage, or increase fossil fuel usage? Or does it prevent, inhibit, or decrease fossil fuel use? This kind of directionality is significant for understanding causal relationships, particularly when the ultimate goal is to be able to influence the systems behavior—this parameter can tell an actor if an intervention will increase the presence of a desire effect, or decrease the presence of an undesired effect.

Because of these two forms, causal direction can be considered as a 2-dimensional variable: given two factors with a causal relationship between them, there is a direction indicating which is the cause and which is the effect, and a direction indicating whether the cause increases or decreases the effect. At the level used in this dissertation, there are thus four different causal directions for a causal relationship between two factors (illustrated in Figure 2.4). For clarity, I will refer to the direction indicating the cause and effect factors as the "target direction" and the direction indicating an increase or a decrease as the "influence direction," with "causal direction" being used to encompass both aspects (the two-dimensional variable).

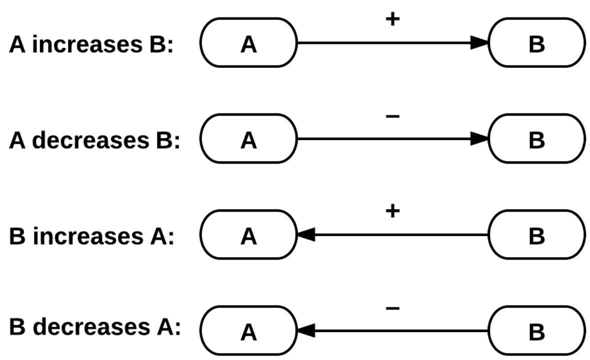


Figure 2.4: The four possible causal directions for a causal relationship.

The concept of influence direction can be considered using a few different interpretations. Plate (2006) frames influence direction as a measure of change correlation: if the changes in the factors are positively correlated, then that causal relationship is said to be "increasing" or "causing an increase" (e.g., an increase in driving causes an increase in fossil fuel usage—the two phenomena are positively correlated). If the changes in the factors are negatively correlated, then that causal relationship is said to be "decreasing" or "causing a decrease" (e.g., an increase in public transportation causes a decrease in driving). Although causation and correlation are distinctly different, the directionality of correlation offers a helpful analog for understanding the influence direction of causation. This view does imply that causal relationships are reciprocal (a fact that may not hold for all models)—an increasing relationship means that an increase in driving leads to (causes) an direct increase in fossil fuel usage *and* that a decrease in driving leads to a direct decrease in fossil fuel usage.

In everyday language, we often use the word "cause" to specifically mean "increase" (in the sense of "promote"), and use words such as "prevent" to mean "decrease" in a causal relationship. Driving *causes* fossil fuel usage, but public transportation prevents driving. Yet a factor can be said to be the "cause" of a preventing relationship—public transportation *causes* people to drive less. The word "cause" is semantically overloaded. Similarly, using terms like a "positive causal relationship" (when positively correlated) or a "negative causal relationship" (when negatively correlated) raises conflicting values-based denotations of whether a relationship is good or bad[9]. To resolve this conflict, Kadaba et al. (2007) use the terms "causal amplification" and "causal dampening" to refer to positively and negatively correlated causal relationships respectively, while Hitchcock (2007) suggests that "it might be perfectly natural to import the language of promotion and prevention"—though he acknowledges potential limitations to this usage for non-monotonic relationships. In this dissertation, I will generally use the terms "increase" and "decrease" to avoid ambiguity in indicating influence direction; for example: "public transportation decreases driving" or "driving increases fossil fuel usage", drawing from Plate's correlation-based interpretation of causal direction.

Overall, in this research the causal direction (both in terms of target direction and influence direction) is the significant parameter for assessing understanding of a causal relationship. Following Plate (2006, 2010), I assess accuracy of understanding primarily on correctly identifying the existence and direction of causal relationships. For my work, a causal relationship is chiefly defined by the causal direction of the relationship between factors; the question of interest is if (for example) public transportation influences fossil fuel usage, and if so whether that relationship is increasing or decreasing.

2.3.2.2 Causal Strength

How much of an increase in fossil fuel usage does driving cause? Does public transportation reduce driving enough to outweigh directly increased fossil fuel usage? Answering these questions requires defining the second most common parameter in discussing causal relationships: the *strength* or magnitude of a relationship. Causal strength is generally a measure of the change in the effect factor relative to the change in the cause factor—if public transportation use changes by a specified about, how much does driving change? At the simplest level, causal strength measures

[9] These are the same kind of denotations that create potential confusion with the idea of an undesirable positive feedback loop and have given rise to the term "vicious cycle".

whether a small change in the cause factor leads to a large change in the effect factor, or vice versa. Such measures are particularly important in complex systems, where relationships may be non-linear (e.g., a small change causes a big effect).

In general, causal strength is defined abstractly as a measure of relative change in the factors[10]—if public transportation usage increases by 10%, by what percentage does driving decrease? Again, change ratios need not be 1-to-1; causality is often disproportionate in complex systems, so that a 2% increase in driving may cause a 5% increase in fossil fuel usage. Causal strength is also often measured probabilistically "in terms of the size of the difference in probability that a causal factor makes" (Hitchcock, 2007), similar to the values in a decision tree or Bayesian network (Pearl, 1988). A probabilistic measurement of causal strength can also potentially encapsulate the stability or regularity of a causal relationship (how often does driving increase fossil fuel usage) inside the total probability. Indeed, total probability can include the aggregates of complex, non-linear causalities, and even the effects of multiple different factors. Causal strength can also be determined using fuzzy logic, as in the field of fuzzy cognitive mapping (Kosko, 1986).

While causal strength is a significant parameter for understanding a causal relationship, the prototype system presented in this dissertation does not consider this parameter in assessing causal understanding. The magnitudes of causal relationships in complex systems and the domain of environmental sustainability in particular are often difficult to measure, and thus may be under significant discussion within the scientific community. Measurements of causal strength depend on the underlying model used to valuate causal relationships and environmental impact, as well as resolving potentially conflicting empirical evidence. As such the accuracy of the causal strength of a relationship is difficult to verify, and thus cannot be properly assessed. For this reason, in this research I follow the example of prior work (Plate, 2006, 2010) and focus on the initial step of assessing people's ability to identify the existence and direction of causal relationships, rather than

[10]This definition assumes that factors have some implicit way of being measured. Linguistically, it means that using count nouns as factors can make the causal strength of relationships more intuitive, as they can be specifically quantified without requiring further units of measurement.

the further refinement of assigning strengths to those relationships.

2.3.2.3 Causal Chains

The example system illustrated in Figure 2.3 shows two ways in which public transportation influences fossil fuel usage. First, public transportation itself *directly* increases fossil fuel usage by consuming fossil fuels in order to run buses and such. Second, public transportation *indirectly* decreases fossil fuel usage by reducing driving, which itself normally increases fossil fuel usage.

This second relationship is an example of a *causal chain*—a sequence of causal relationships between multiple factors. Perhaps because of people's tendency to focus on a single event at a time as well to explain the world using simple narratives (see above), they often think about causality in terms of causal chains—even when such chains are not the best models of a causal system (Assaraf and Orion, 2010; Plate, 2006; Raia, 2005). Indeed, human understanding of causality allows us to chain together causal relationships into "causal paths" (Tomlinson and Black, 2011a) and perceive a path as a single causal relationship. This process implies that causal relationships are multiplicative—chaining an increasing relationship and a decreasing relationship makes for an overall decreasing relationship, as with public transportation decreasing fossil fuel usage via decreasing driving.

Thus viewing causal chains as a single causal relationship involves a "zooming out" process (Lyons et al., 2012), in which a set of factors are either abstracted into a single factor, or the chain of relationships between ultimate cause and effect factors are abstracted into a single relationship. Alternatively, a single causal relationship can be "zoomed in" and broken up into a chain of relationships—for example: driving increases the use of internal combustion engines, which increases fossil fuel usage. Indeed, almost all considerations of causality involve selecting the "zoom level" in some way, if only in choosing which factors to consider as part of the system—for example, there are numerous intricate chemical reactions that occur as part GHG emissions causing climate change, yet these are often abstracted away. One can almost always be more detailed and include more intermediate steps in explaining a causal relationship. For this reason, picking the level of detail and which factors are considered (and which are obscured) is an important step in modeling complex systems, and can have a significant influence in how people think about causal relationships. Long causal chains (such as in Figure 2.5) can become difficult to parse, but can also reveal new, non-intuitive relationships between distantly related factors. Helping people to identify and process these long, non-intuitive chains is one of the goals of the Causality Project (Tomlinson and Black, 2011a,b), a broader research effort that encompasses some of this dissertation work and that is detailed later in this chapter.

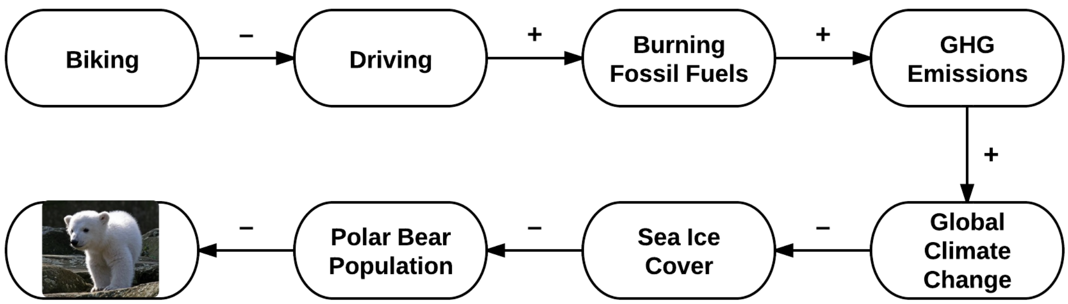


Figure 2.5: An extended causal chain (not including any other causal relationships) that shows how biking leads to more adorable polar bear cubs.

Given multiple separate causal paths or chains between two factors, one may want to resolve or combine them into a single relationship: overall, does public transportation increase or decrease fossil fuels? These individual paths—referred to as component effects (Hitchcock, 2001) or path-specific effects (Pearl, 2001)—all have some influence on the ultimate change in the paths' shared effect factor. It may be possible to resolve these multiple paths by aggregating their causal strengths; if public transportation directly increases fossil fuel usage by 2%, but indirectly decreases fossil fuel usage by 5%, we might say public transportation net decreases fossil fuel usage by 3%. However, in this research I am more interested in the systems thinking skill of identifying the variety of causal paths, rather than attempting to combine them all into a single simplified model—how people "zoom in" on causal chains rather than how they "zoom out." In complex systems, the presence of emergent effects mean that such aggregations are rarely straight-forward or may not be calculable through simple additive methods. Sustainability rarely has a single answer to "which is a greater effect"; indeed, such aggregation may obscure the presence of other significant factors or side effects.

2.3.2.4 Other Causal Parameters

There are a variety of other parameters that can be considered when discussing causal relationships. For example, we could consider the presence of *causal multiplicity* or joint effects—when a causal relationship occurs because of the combination of two or more factors. One factor may not have a causal effect on another by itself, but does when combined with a third factor as a catalyst; for example, driving may only increase fossil fuel usage when combined with an internal combustion engine. Drawing on Plate's (2006; 2010) methods, my assessment of causal understanding does not include this detail, though factors can be defined so that they act as aggregates of multiple factors (e.g., "driving diesel cars" could be a factor that combines "driving" and "diesel cars").

It is also possible to think about causal relationships in terms of the influence of alternative factors. For example, when considering the relationship between public transportation and fossil fuel usage, one may ask "public transportation as opposed to what?" While the "public transportation decreases driving" relationship in a way captures this alternative (taking public transportation as opposed to driving), it maybe possible to model a system of causal relationships as including choices of alternatives. For example, a system could model the difference in effects on fossil fuel usage between driving, taking public transportation, biking, walking, or just avoiding traveling at all. The existence and effects of alternate actions are particularly important when considering interventions to increase sustainability (see Ross, 2011; Ross and Tomlinson, 2011); such a framing could be useful for future assessments of causal understanding in complex systems. However, this level of detail can complicate the reification of causal models needed for assessment of causal understanding, and thus is not used in this dissertation.

2.4 Causality and Cognitive Mapping

As demonstrated in the figures in the preceding sections of this chapter, the relationships that make up causal systems can be diagrammed as *causal maps* (Axelrod, 1976; Eden et al., 1992; Plate, 2006). Causal maps are a form of cognitive mapping (Kearney and Kaplan, 1997; Tolman, 1948): they are concept maps (Novak and Cañas, 2008) that aim to visually represent a person or group's beliefs about the causal relationships of a system—causal mapping is thus the process of representing a person's beliefs in the form of a causal map. Causal maps are most commonly depicted as directed graphs, where vertices/nodes are the factors making up the system, and the edges/arcs are the causal relationships between the factors. An edge is drawn as an arrow pointing in the target direction from the cause factor to the effect factor, and often labeled with a plus or minus sign to indicate an increasing or decreasing relationship. Thus in Figure 2.3, "public transportation", "driving", and "fossil fuel usage" are the vertices of the graph, with labeled edges between them marking the causal relationships. While many causal maps are constructed as acyclic directed graphs (for better use in calculating probabilities or making decisions), causal maps of complex systems include cycles to indicate and help with discovery of feedback loops. As representations, causal maps need not include all existing causal relationships or even be accurate representations of the "true" complex system; causal maps are intended to model a set of beliefs about a system as reified through the construction process.

The field of cognitive mapping (in which causal mapping can be included) derives in a large part from the work of Edward Tolman, who suggested that rats constructed mental maps of a maze in order to navigate it, rather than simply relying on responses to stimuli (Tolman, 1948). While cognitive mapping has been extensively used in considering how people and animals navigate spatial environments, the notion of a cognitive map has also been extended to conceptual environments—how people organize concepts, processes, or situations rather than landmarks (Kearney and Kaplan, 1997). Thus Kearney and Kaplan describe cognitive maps as "hypothesized knowledge structures embodying people's assumptions, beliefs, 'facts,' and misconceptions about the world" (Kearney and Kaplan, 1997). Robert Axelrod introduced causal mapping as a form of cognitive mapping in the context of political science (Axelrod, 1976), and this method was soon picked up and used extensively in the field of management studies (e.g., Eden et al., 1992; Fahey and Narayanan, 1989; Markóczy and Goldberg, 1995; Scavarda et al., 2006). In this field, causal maps are usually constructed from the beliefs of managers and other members of an organization, either to search for conflicting understandings by comparing causal maps or to identify key causes and interventions for problems in the organization's operations. Causal maps provide a method of soliciting knowledge and beliefs about complex systems from organizations.

There are a number of methods for analyzing and interpreting causal maps, particularly using quantitative analysis. For example, (Eden et al., 1992) provides a variety of measures for determining the complexity of a map, ranging from a simple count of the number of vertices and edges, to the calculation of the ratio between them, to using domain analysis to consider the number of "in-arrows" and "out-arrows" and measure graph branching. Along these lines, researchers often consider causal maps using techniques from graph theory. Maps are converted into adjacency matrices, which can then be analyzed mathematically (see Markóczy and Goldberg, 1995; Özesmi and Özesmi, 2004, for examples) to determine measures such as density and factor centrality. More emergent properties of causal maps, such as factor hierarchies, can also be detected through clustering methods (Eden et al., 1992; Markóczy and Goldberg, 1995). Based on such clustering, Markóczy and Goldberg (1995) present methods for computing distances between causal maps, thereby allowing maps produced by different people to be compared. Alternatively, Plate (2006, 2010) computes similarity scores between causal maps using an agreement rubric, described in more detail below.

These analysis methods have enabled cognitive and causal mapping to spread beyond the field of management studies, such as to the domain of environmental sustainability. For example, Özesmi and Özesmi (e.g., 2003; 2004) have used causal mapping in the form of a type of fuzzy cognitive maps to develop ecosystem management plans for a lake and wetland ecosystems in Turkey. Similarly, Kearney et al. (1999) used cognitive maps to study stakeholder perceptions in forest management, finding in particular that stakeholders tended to view one another through unsubstantiated stereotypes (e.g., as the greedy timber industry). Kearney and Kaplan have also considered how interpreting cognitive maps can support environmental education:

> "Identifying these gaps in environmental knowledge is indeed important. An environmental educator, for instance, hoping to design an effective intervention, must first discover the types of information people need. By itself, however, the identification of knowledge gaps does not provide sufficient direction for devising education and communication strategies. Another important factor to consider is an individual's existing mental model, or 'cognitive map,' of the issue" (Kearney and Kaplan, 1997).

In this way, the authors describe cognitive mapping as a tool for assessing and communicating knowledge about environmental systems. Indeed, concept maps such as causal maps are increasingly being used for assessment in a variety of pedagogical contexts (e.g., McClure et al., 1999; Novak, 2010). Combining these domains, Plate has used cognitive causal mapping to assess the effects of systems thinking pedagogies on students' understanding of non-linear causal structures in environmental systems (Plate, 2006, 2010). Plate analyzed causal maps produced by students based on a "web-like causality index" (how much the map branches from being a single linear path), the presence of causal loops, and accuracy as compared to expert models. Overall, these examples demonstrate how causal mapping can be used as a methodology for measuring causal understanding of topics in environmental sustainability.

A number of methods for constructing causal maps have been developed and validated (see below) as producing indications of how people understand systems. Thus causal mapping is an established and appropriate technique for assessing causal understanding, providing a basis of prior research for this dissertation. However, like any methodology, cognitive and causal mapping have a number of limitations. In particular, scholars (see Nicolini, 1999) have questioned the validity of the method—how well can causal mapping be said to reveal people's understanding of a causal system? For one, causal maps are representations created by fallible people—they may not match any "truth" about the system described. As Markóczy and Goldberg say: "If we construct CMs [cognitive maps] from individual's stated beliefs, then a CM is only about her/his system of stated beliefs" (Markóczy and Goldberg, 1995); such maps are subject to all the same biases as the subject producing the map. Furthermore, cognitive maps are constructed "representations of representations" (Nicolini, 1999)—the maps produced may not actually match what is in people's heads. Thus there may be questions about the true validity of cognitive maps as a method for revealing causal understanding.

For this dissertation, I adopt the view that causal maps can represent *an* interpretation of a person's understanding of a causal system, though not necessary the only (or most accurate) interpretation. Causal maps are a representation of a system, and so a reification of this representation (whether in a diagram of in a textual explanation such as a traditional exam question or paper) is also a reification of how a person understands the system. Indeed, this assumption forms the basis for much of pedagogical assessment at the understanding level (see Bloom et al., 1956): if a person can explain and reify a concept, then he or she must understand it. While causal maps may not be direct representations of what is in people's heads, such maps should be at least consistent with understanding[11]. This work follows from the assumption that cognitive mapping can thus be an adequately representative technique for measuring causal understanding. Moreover, this thesis explores the possibility that the process of creating these maps can be informative about the *ways* in which maps may succeed or fail at matching a person's actual understanding.

This dissertation follows Plate (2006) in that it measures the accuracy of causal maps in comparison to the maps constructed by groups of experts. Being constructed by people with potentially flawed understanding, such maps are not guaranteed to be accurate representations of reality (i.e., scientific reality). Nevertheless, this research and the assessment strategy it entails is based on the

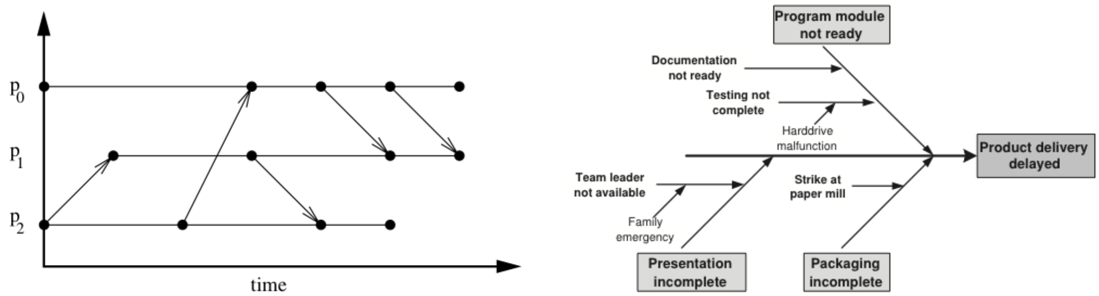
[11]Though note that in the interactive system presented in this dissertation, subjects may not be consciously aware that they are creating a causal map. Nevertheless, I assume that such constructed maps continue to represent understanding due to the design of the interactive system; see Chapter 4.

assumption that expert maps reflect a well-informed and "best possible" representation of reality, which in the end is filtered through human perception. Causal maps may not represent the truth of a system—indeed, the whole point of assessing understanding of systems is that people may have things wrong. Given these limitations, comparing causal maps to those representing the perceptions of a collection of trained experts can provide a useful assessment of causal understanding.

2.4.1 Visualizing Causality

Though I have already described the basic method of diagramming causal maps (as directed concept maps), there has also been a significant amount of previous work in determining how best to visually represent causal relationships. Such prior research provides groundwork for this dissertation, with its reliance on visual presentation to support users in performing causal mapping. Indeed, effectively visualizing causality is one of the top unsolved problems in information visualization (Chen, 2005).

Outside of the field of cognitive mapping, causality has traditionally been visualized either through Hasse (e.g., time-space, Figure 2.6a) diagrams or Ishikawa (e.g., fishbone, Figure 2.6b) diagrams. Hasse diagrams are used to show the causal relationships between processes over time while Ishikawa diagrams are used to categorizing causal factors, including factors with indirect influence. Causal concept maps, on the other hand, focus on showing the "web-like" variety of causal relationships in a system, as well as the relative directions and strengths of those relationships. Thus the method of visualizing causality depends in part upon what elements of causality one wishes to present—the temporal causal relationships, the variety of causal factors influencing a single effect, or the interconnected web of causal relationships. This dissertation considers the last, and thus is based on the concept map form of visualization.



(a) Hasse **(b) Ishikawa**

Figure 2.6: Examples of Hasse (a) and an Ishikawa (b) diagrams. Images from Elmqvist and Tsigas (2004) and Kadaba et al. (2009), respectively.

In order to visualize causality effectively, a causal concept map needs to illustrate each parameter of a causal relationship—particularly the direction and the strength of the relationship. The target direction of a causal relationship is regularly indicated with an arrow pointing from the cause to the effect, but visualizing the influence direction and causal strength can be more intricate. Zapata-Rivera et al. (1999)[12] and Kadaba et al. (2007, 2009) describe a number of visual techniques for displaying the influence direction and strengths of causal relationships:

1. **Color:** Color can be used to indicate the influence direction and/or strength of a relationship (e.g., red edges are increasing and blue edges are decreasing; brighter colors are stronger relationships). Zapata-Rivera et al. assign effect factors a color based on the color of their causes—if the cause factors are blue and red, then the effect factor is colored magenta. This allows color to indicate the indirect influences of a factor (see also Elmqvist, 2004; Elmqvist and Tsigas, 2004).
2. **Size:** The thickness of an edge can be used to indicate causal strength, or nodes can be sized based on their marginal probability and influence on other factors (adopting a Bayesian view (Gopnik et al., 2004) of causality).
3. **Position:** Nodes can be positioned such that the distance between them reflects the strength of the causal relationship—the greater the strength, the closer the nodes are to one another.

[12] Note that Zapata-Rivera et al. consider causal maps in the context of an acyclic Bayesian network, in which causal strength is considered as the combined probability of factors, rather than magnitude of influence.

4. **Iconography:** A standard method for indicating the parameters of a causal relationship is to use icons or glyphs to label the nodes and edges: usually a plus or minus sign attached to the edge. Kadaba et al. (2007, 2009) use such glyphs attached to nodes, along with a small bar graph to summarize relative strengths. Visual techniques can also be applied to the icons: for example, bigger glyphs can represent stronger relationships.

These techniques can of course be combined to represent multiple parameters or to overload the visualization. For example, color could be used to indicate influence direction, while size and icons can both reflect causal strength.

All of these techniques have potential limitations when used for displaying the web of causal relationships found in complex systems such as environmental sustainability. Colors can have culture- and value-dependent connotations (Barber and Badre, 1998), particularly when considered within a sustainability content (e.g., the color green's frequent association with "sustainable" may complicate the visualization). Creating icons that are easy to intuit and look aesthetically pleasing can be a challenge, and displaying too many icons can make the map cluttered and difficult to read. Adjusting size has similar limitations—how does one make size changes significant enough to be noticeable at a glance, but still leave room for drawing a potentially large number of nodes and edges? Positioning factors that are connected by a large number of relationships can be algorithmically complex, and may not support cyclical relationships and feedback loops (e.g., when there is a strong A-causes-B relationship, but also a weaker B-causes-A relationship). Indeed, positioning nodes to reduce edge crossing and make the map more readable, such as through force-directed layout algorithms (Kobourov, 2007), is a valuable technique in its own right.

Another increasingly studied technique for visualizing causality is animation. Animation is a "perceptually efficient display dimension (Bartram, 1997)" (Yao, 2008) that lends itself to displaying change over time—a common aspect of causality to visualize, such as in Hasse diagrams. Animation techniques often draw from Michotte's experiments looking at how people infer causality from visual events (Michotte, 1963). For example, Michotte found that if people see a moving object L stop in close proximity to a second object T which then begins moving, they perceive a causal relationship: L struck T and caused it to move. This effect is known as *launching* (see Figure 2.7). While there has been some criticism of the specifics of Michotte's findings in terms of timing and speed parameters (see Yao, 2008), his basic claim that we infer causality from visual stimuli has been upheld, and as such has been used to visualize causality through animation. For example, Kadaba et al. (2007, 2009) use animated "bullets" (which can themselves be glyphs) moving along the edges in the target direction; when these bullets hit the effect factor node, that node changes in size based on the influence direction of the relationship. However, this method focuses on the causal relationships leading to a single ultimate effect (similar to in an Ishikawa diagram), rather than a full web of causality. Another implementation of animated causality, Growing Polygons (Elmqvist, 2004; Elmqvist and Tsigas, 2004), visualizes the interrelations among processes or events over time by having color-coded polygons representing the processes increase in size over time as they are influenced by one another. Although this technique allows for more complex interrelationships, it continuous to focus on visualizing the change over time while my research focuses on visualizing the complexity of the causal web.

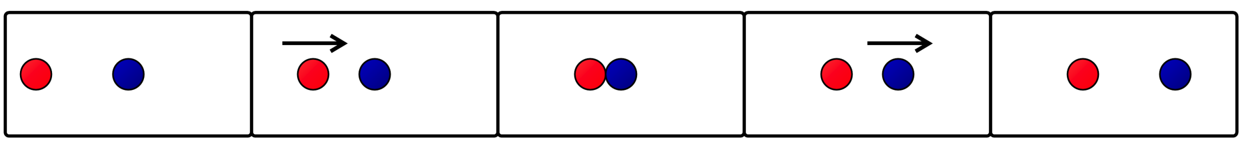


Figure 2.7: Michotte's launching effect. For animated demonstrations of this effect, see `http://www.yale.edu/perception/Brian/demos/causality.html`.

In developing the system presented in this dissertation, I have considered all of these techniques for supporting the visual construction of causal maps—with the further requirement of supporting game-based information and aesthetics. The details of the developed system's visual appearance are discussed in Chapter 4.

2.4.2 Constructing Causal Maps

I have discussed some of the uses of causal maps as well as what they look like, but where do these maps come from? As stated previously, causal maps are constructed representations of people or organization's beliefs about a system. The form and content of a causal map is highly dependent on the method used to construct it—as Markóczy and Goldberg say: "The meaning of a causal map is not only a function of the map itself, but of the way in which it is elicited" (Markóczy and Goldberg, 1995).

The oldest and most common method of eliciting causal maps from people is to have researchers transcribe user interviews into causal maps, following the method originally used by Axelrod (see Axelrod, 1976). In these interviews "qualitative and open-ended questions are posed to experts to obtain raw data in the form of narratives" (Scavarda et al., 2006). Interviews are often multi-stage: early interviews identify important factors in the system, and subsequent interviews focus on the causal relationships between those factors (see e.g., Jenkins and Johnson, 1997). There may also be a stage in which produced maps are validated with the interviewees. Once the raw data from these interviews are collected, they need to be coded into causal maps for analysis—researchers try to pull out sentences that suggest causality (looking for keywords such as "produces", "causes", "affects", etc) to use in constructing the map. This requires a number of subjective measures, as Plate quotes:

> "Axelrod concedes that such a procedure 'requires a large number of subtle coding decisions,' but explains that 'after more than three years of work, the coding rules [which comprise a forty-page appendix to the text] have reached a state of precision such that the intercoder reliability is fully compatible with the accepted standards of good quantitative work in the social sciences' (1976)" (Plate, 2006, brackets from Plate).

Thus while historically common, the interview method of eliciting causal maps is time and effort intensive, as well as easily subject to coder and participant bias. Researchers have looked to reduce this bias, such as through systematic coding of existing documents (e.g., Carley and Palmquist, 1992; Nadkarni and Narayanan, 2005). Other researchers have considered other variant methods for producing causal maps. For example, Özesmi and Özesmi (2004) cite the use of questionnaires (in place of interviews) and data extraction as other options, while Scavarda et al. (2006) present a methodology for having groups of experts collectively generate causal maps. Furthermore, the reliance of such methods on soliciting "narratives" of causality may be in conflict with the systems thinking requirements needed to understand complex systems (see above).

One common alternate methodology for eliciting causal maps is the Self-Q Interview (Bougon 1983; described in Nicolini 1999; Plate 2006). In this method, participants first generate a list of questions about a system under examination. The researcher converts these questions into a list of factors (using original wording), which the participant then validates by selecting the top 10 or 11 to include in the map. Once these factors are chosen, the participant then goes through each pair of factors and decides whether there is a causal relationship between them (and if so, what its direction and possibly strength are). In effect, participants create their own causal maps. This method has been repeatedly used and validated (Nicolini, 1999)—Eden et al. call it one of the "few well-developed methods for the elicitation of cause maps" (Eden et al., 1992). As such, this technique offers a validated, systematic methodology for enabling participants to construct their own causal maps, removing from the researcher the need for extensive coding as well as the potential to introduce bias. Indeed, the 3CM method (Kearney and Kaplan, 1997) for cognitive mapping that Plate (2006, 2010) bases his work on is very similar to the Self-Q technique. And as this dissertation draws heavily from Plate, the Self-Q technique is something of a granduncle to my own methodology, and many of its assumptions are continued in this work.

The Causality Project (Tomlinson and Black, 2011a,b) is another system for constructing causal maps that informs my research (a system that was developed by my research group and to which I contributed). The Causality Project is an online platform[13] for collecting and visualizing causal knowledge that is distributed not only among experts, but also among everyday users who interact

[13]The Causality Project platform can be found at `thiscausesthat.org`.

with the complex systems that make up human civilizations and the natural world. In effect, this project seeks to compile a massive database of causal relationships between factors in the broad domain of environmental sustainability and beyond. In this platform, each causal factor included in the database can be displayed on a unique web page (similar and building on the "article page" format of Wikipedia), as can each individual causal relationship. Users are able to browse and explore these factors and relationships, as well as add new items to the database (again, similar to authoring a new article on Wikipedia)—a reputation management system helps to encourage such contributions. Furthermore, the Causality Project platform includes a visualization component that dynamically displays the relationships stored in the database as a causal concept map. This visualization allows users to easily browse and explore the entire causal web included in the database, providing the opportunity to discover causal relationships of which they previously were unaware. In addition, a significant focus of the Causality Project is on the identification and consideration of linear causal chains. The system can compute the causal chain between any two factors in the database using a path-finding algorithms—users can thus search for the causal link between any factors they are interested in, in order to "understand more fully how the environmental issues in our world are connected" (Tomlinson and Black, 2011a). In these ways, the Causality Project aims to solicit the collective causal knowledge of a large number of users, combing this knowledge into a single, comprehensive causal map that can be explored and searched for linear causal chains.

It is worth noting that all of these methodologies and systems suffer from the fact that they ultimately produce representations of representations, and thus may not actually represent the causal models that exist in people's heads, much less the causal relationships actually at play in the physical world. As Axelrod says: "The cognitive mapping approach is, of course, in no better (or worse) position in this regard than any other procedure that relies on a person's conscious and monitored linguistic behavior to make inferences about his or her beliefs" (Axelrod 1976; cited in Markóczy and Goldberg 1995). In spite of this limitation, as previously argued, cognitive causal mapping still represents the best and most commonly used approach to measuring and recording people's understanding of a causal system.

Although there has been extensive research into the generation and analysis of causal maps (of which this chapter gives only an overview), all the examples found in my survey focus on analyzing a completed causal map, whatever method used to create it. Map creation methodologies include a number of steps, and yet the *process* of creating these maps has little or no part to play in the analysis. For example, analysis does not consider whether interview participants revise their statements of causality as they weave narratives of the system, or whether they hesitate about assigning direction to a relationship in the Self-Q technique. Causal maps are treated as produced artifacts, with the production process discarded from consideration. This exclusion is understandable given logistical limitations—the process of producing a causal map could itself be seen as a complex system; recording and analyzing this process is generally not feasible. In this dissertation, I begin to address this absence with a computational system in which users can build their own causal maps, automatically collecting data on the map construction process to be able to analyze this behavior and how it reflects on causal understanding.

2.5 Conclusion

In this chapter I have surveyed previous work surrounding the topics of causality and causal mapping. I have attempted to scope the meaning of "causality" for the purposes of this research, and provided a set of semantics that can be used to discuss the relationship between cause and effect. I have also discussed this understanding of causality in the context of complex systems, and considered some of the ways that causal understanding and systems thinking may or may not be linked. Finally, I explained the concept of cognitive causal mapping that forms the basis of this research and the process by which I will assess the accuracy of causal understanding.

The next chapter continues my review of prior and related work by considering how the domain of computer games (as a form of interactive computer system) may offer support for the process of building causal maps in order to document and assess knowledge of causal relationships.

Chapter 3. Background and Related Work: Games

Continuing my discussion of related work, in this chapter I ground my research's consideration of games as a form of interactive system for use in assessing causal understanding. I survey previous research into games, particularly research in the domain of serious games. I examine how games have been positioned as effective at supporting education, comparing this framing to the use of game dynamics in assessment. I conclude with an overview of techniques developed in the field of gamification, and how such techniques may be capable of supporting games for assessing causal understanding.

3.1 Games and Play

In everyday English, the term "game" (and similarly, "play") is used in a vast variety of ways—as Parlett says: "The word is used for so many different activities that it is not worth insisting on any proposed definition. All in all, it is a slippery lexicological customer, with many friends and relations in a wide variety of fields" (Parlett, 1999; quoted in Salen and Zimmerman, 2003). Nevertheless, academics and scholars have long argued about definitions and meanings to help ground our understanding of game systems and activities. One of the earliest foundational definitions comes from Johann Huizinga, who describes play as "a free activity standing quite consciously outside 'ordinary' life as being 'not serious,' but at the same time absorbing the player intensely and utterly" (Huizinga, 1955). This description emphasizes the "non-serious" and playful nature of games, along with their ability to engage and absorb a player. Salen and Zimmerman attempt to unify this and other previous descriptions, arriving at the definition: "A game is a system in which players engage in an artificial conflict, defined by rules, that results in a quantifiable outcome" (Salen and Zimmerman, 2003). Juul also synthesizes a similar but more detailed definition: "A game is a rule-based formal system with a variable and quantifiable outcome, where different outcomes are assigned different values, the player exerts effort in order to influence the outcome, the player feels attached to the outcome, and the consequences of the activity are optional and negotiable" (Juul, 2003, 2005). These definitions point out the structured, rules-based nature of most games, as well as the requirement for a conflict or challenge upon which the outcome is dependent—it isn't a game if there is no chance of losing! Indeed, philosopher Bernard Suits offers one of my favorite definitions: "playing a game is the voluntary attempt to overcome unnecessary obstacles" (Suits, 1978).

So what should we take away from these definitions? These definitions emphasize how games—and video games[1] in particular—are structured systems of conflict. Games are structured by their rules. In almost any game, the rules define a set of "moves" (e.g., Lindley, 2005) that players can take, specifying what actions players are allowed to perform within the formal system of the game (e.g., what is a legal move in chess, or how far you can jump in *Super Mario Bros.*). Rules-defined moves have meaning within the what Huizinga calls the *magic circle*: "the arena, the card-table... forbidden spots, isolated, hedged round, hallowed, within which special rules obtain" (Huizinga, 1955). In the magic circle or performance frame (Bateson, 1955) of the game, rules provide actions with particular meanings, so that the player actions are interpreted within the context of the game (in addition to—or potentially in conflict with—any meanings existent outside of the game). In this way, like other computational systems, games can be defined by their procedurality (Bogost, 2007;

[1] I use the term "video game" to refer to digital interactive games, whether played on a general computer, a specialized console connected to a television, or a mobile device. And while my dissertation focuses on video games as a form of interactive system, this discussion of the theory of games applies equally well to non-digital games such as card and board games (Zagal et al., 2006).

Juul, 2005; Murray, 1997): the rules of a game define the procedure of how a player interacts with and experiences that game. As Juul says: "the rules of a game add meaning and enable actions by setting up differences between potential moves and events" (Juul, 2005). It is the way that rules structure a game that makes playing that game meaningful (Salen and Zimmerman, 2003).

Game rules are most commonly understood as the formalized structures found in the instructions or allowed by the computer game engine, but rules can also be informally established, verbally or non-verbally, through the social contract established by the players agreeing to play (Sniderman, 1999). For example, players might place limits on what characters or moves are fair to use in a fighting game such as *Super Smash Bros.* (Hung, 2011; Jakobsson, 2007), or even design their own games and rules within an existing system (Parker, 2008). Indeed, this distinction between formal and informal rules follows from Caillois' (1961) distinction between *paidia* (the unstructured, spontaneous free play often performed by young children) and *ludus* (games with structured, specific rules)[2]. Games are generally structured by a complex mix of rules as they are written down, rules as they are understood by the players, and rules as they are actually enacted by the players (Wilson, 2011, see e.g.,). The designed rules of a game shape and influence how players play a game, but players can also choose which rules are followed.

This distinction highlights two different views of what it means for an activity to be a game: games are marked by formal systems, or games are marked by social agreement. Many of the components of the definitions listed at the star of the section frame games as formal systems: a game is a collection of rules that structure actor behavior through particular restrictions. This systemic view is closely related to that adopted in the field of game theory, which studies "games" as being simply interactions between decision-making agents. The important component of a game is how the rules define interaction with a system; indeed, from this perspective any system in which actors make conflicting decisions can be considered a "game."

[2]The exact implications of the Greek words *paidia* and *ludus* for modern games are open to some debate. For example, (Frasca, 2001) uses *ludus* to refer to games that have a definite winner and loser, and *paidia* to games that do not. And although *paidia* and *ludus* are sometimes taken to mean "play" and "game" respectively, I believe that any game can have both *paideic* and *ludic* elements.

However, the concept of a game can also be viewed through the lens that borrows more from considerations of play and social interaction (e.g., Huizinga, 1955): a game is identified not by its structured system of conflicting decisions, but because it is socially positioned *as* a game. The fact that a game occupies a space and semantic meaning that is set apart (i.e., the magic circle) allows it to be considered a game. Indeed, from this view, an activity can be considered a game simply by calling it such—a game exists whenever the actors agree that it should be treated as one. If an activity is called a game—and the player believes it is a game and so approaches it with a "playful attitude" (Schell, 2008)—then that activity is in fact a game. It is this social view (rather than a systemic view) that emphasizes the requirement that a game be "fun," and thus can be seen as matching with a more every-day definition of "game." An activity being a game is less a function of its systemic components, but of the interaction's social framing.

These two views focus on dramatically different interpretations of a game. The systemic view focuses more on the interactions of the actors with the rules of the system, while social framing focuses more on the social context in which the activity was positioned. In this dissertation, I draw from both interpretations. I use a systemic view to position games as a particular form of interactive system, and consider how explicitly framing system usage as driven by a particular set of rules (e.g., the rules and goals required to "win" the game) may influence interaction. Moreover, this view of games as distinct forms of complex systems motivates some of the application in the domain of education (see below), which I build upon in this research. On the other hand, much of my consideration of how game dynamics may be used to support assessment relies on techniques that consider games more as social interactions than formal systems. Indeed, I consider the process of framing an existing interactive system as a game to be primarily a social act; I claim that by calling a system a game (and providing it with trappings that signal to the user that the system should be approached playfully), a system is in fact made into a game. This research considers the effectiveness of such demarcation for use in assessment: what elements signal that a system is a game (rather than an assessment), and how might these elements support assessment processes? Both views are useful for considering what it means to make a system into a game, though the social consideration may have more significance for considering user interaction with the proposed system and techniques. Overall, I consider games to be systems that have been socially determined to be games by the participants as such—however, there may be common systemic elements (e.g., points or levels, computer platforms used, etc.) that semiotically work to suggest that a particular designed system should be considered as a game.

Indeed, most systems that are socially designated as games are those that have been explicitly *designed* to be approached and viewed as games. There are a vast number of books, papers, and annually-held conferences about the process of game design (for prominent treatments, see Fullerton et al., 2008; Salen and Zimmerman, 2003; Schell, 2008). Perhaps in reflection of the slipperiness of achieving social agreement on what is considered a game, one prominent theme in many of these books is the ability and need to consider designed games from a variety of perspectives. For example, Salen and Zimmerman (2003) offer a number of schemas for viewing and creating games, such as Games as Emergent Systems and Games as the Play of Meaning (considering games as representations of the real world). Schell (2008) also provides a list of 100 lenses that are applicable to game design—and covering similar topics (e.g., he includes a Lens of Emergence). This need for multiple, successive viewpoints makes *iterative design* (Fullerton et al., 2008; Salen and Zimmerman, 2003) an appropriate a method for creating games, emphasizing prototyping, playtesting, and continual refinement of a game to move it closer to an experience that all players will agree can be considered a (good) game. Yet despite this plethora of discussion and theory, game design remains a skill and an art—exceedingly difficult to perform and master. Indeed, Carl Jung said: "One of the most difficult tasks people can perform, however much others may despise it, is the invention of good games" (quoted in Fullerton et al., 2008). Although this dissertation does present the design of an interactive game system, my focus and contributions with this research is on how particular designed elements that signal a system as a game (the game framing) may influence assessment, rather than on a particular process for designing such games.

One of the characteristics that leads people to socially designate an activity as a game is whether the activity is in some way playful or fun[3]. According to theories of motivation (e.g., Ryan and Deci, 2000), a game is thus an *intrinsically* motivated activity—that is, a game is inherently enjoyable and makes playing its own reward. This is in contrast to *extrinsically* motivated activities that lead to rewards separate from the task (most commonly monetary, but also the acquisition of skills or knowledge that can be applied in the future). Note that like any activity, even systems that are socially designated as games can involve any combination of intrinsic and extrinsic motivations: a person may enjoy playing *StarCraft*, but be extrinsically motivated to win a national tournament (Cheung and Huang, 2011; Hutchins, 2008), or a professional baseball player may play the game only for the money—it in effect has become a job, and might no longer be considered a game if it is not considered such by the player[4]. Indeed, the split between intrinsic and extrinsic motivation is one way of describing an activity as either "play" or "work"—a distinction that may directly correspond with whether an activity is a game or not.

On the other hand, it is worth noting that elements that are often perceived as "game-like" (that is, elements that may signal an activity's status-as-game to a user) such as points, high scores, and other specified achievements (Hamari and Eranti, 2011) can act as extrinsic rewards, offering some value outside that of the game activity itself. Such rewards can potentially "crowd-out" the intrinsic value of the activity, causing what is nominally a game to instead become a "work"-like grind (Jakobsson, 2011). The line between work and play—between extrinsic and intrinsic motivation—is often permeable: work can sometimes be fun, and games can sometimes feel like work. Indeed, Yee suggests that many games (and online multi-player games in particular) make the line between work and play indistinguishable: "The point remains however that video games are inherently work platforms that train us to become better workers. And the work being performed in video games is increasingly similar to actual work in business corporations" (Yee, 2006b). The fact that video games are still perceived as enjoyable (and as games) despite the "grind" (Rettberg, 2008) and keep

[3]Because a person's viewpoint can influence an interaction, activities may be seen as fun because they *are* games in the same way that activities may be seen as games because they are fun

[4]Though baseball as a formal system could still be considered a game from a systemic perspective that does not rely on the subjective views of the participants and/or observers.

us coming back for more speaks to the strength of the social framing that positions them as games: by calling an activity a game, players continue to approach it as such even if the activity feels like work. Despite the often work-like activities involved in video games, they still exist within the socially signaled (and thus perceived as non-serious) "separate" space of the magic circle, where actions take on game-based meanings that help make them intrinsically motivating.

Nevertheless, establishing and maintaining this social framing (the magic circle) can be difficult for games with goals beyond simple entertainment—such as a game to assess causal understanding. Such games may run the risk of breaking the magic circle, and hence becoming perceived and interacted with as work rather than the motivating and engaging play of video games.

3.1.1 Engagement in Games

The view of games as being socially determined activities in which people volunteer to engage places significant emphasis on the intrinsic motivation that arises from framing something as a game—the fact that games are fun and engaging experiences. Indeed, video games [5] demonstrate exceptional levels of engagement—what Turkle (1984) calls *holding power—such that people volunteer to engage with these activities for hours on end.* For example, an empirical study by Ducheneaut et al. (2006) found that players in the online game *World of Warcraft* spent an average 10.2 hours a week playing that single game; surveys by Yee (2006a) estimate as high as an average of 22 hours per week playing this game. Yet even the lower estimate suggests that in the first 8 months the game was available, 15% of players had each reached "an accumulated play time of 15.5 days - a total of 47 8-hour work days, or roughly two full months of work days" (Ducheneaut et al., 2006). The single game of *World of Warcraft* currently has more than 10 million subscribers worldwide, demonstrating the pervasiveness of this level of engagement. Overall, consumer research group

[5] Note that this chapter's discussion of games' properties and benefits applies primarily to well-designed, "good" games (see Gee, 2003b)—the kind of games that game designers strive to make. As any activity can be considered a game if it is called as such, not all game systems demonstrate these traits. There are poorly designed games (or activities that make poor games). See also subsubsection 3.1.1.1

NDP estimates that top gamers spend 16 to 18 hours a week playing games (NDP Group, 2011), while McGonigal (2011) suggests that the average young American spends as much time playing video games (around 10,000 hours before age 21) as they do attending school.

Moreover, this engagement with video games also drives engagement in *other* activities. For example, players may write highly detailed descriptions and walkthroughs for their favorite games, which "often run to 70 or more single-spaced pages" (Gee, 2003a). The player-created *World of Warcraft* wiki is the second largest English-language wiki after Wikipedia (McGonigal, 2011). Gamers write blog posts and produce machinima (Lowood, 2007), and otherwise use games as a springboard into engaging in forms of participatory culture (Jenkins, 2006, 2009). All this time spent playing and producing media (Pearce, 2006) about games points to games' holding power—their ability to draw in and engage people, to keep players involved and participating in activities that may be only loosely related to the game.

Furthermore, the level of engagement and participation found in games may be highly desirable for other activities or interactive systems—for example, educators may wonder how to get students as engaged in mastering school content as they are in mastering video game systems. Thus the question is raised: how can the engagement of games be harnessed to drive participation in other activities? In order to utilize the engagement of games for other purposes, we need to first consider what causes games to be engaging in the first place. Does this engagement arise from the interaction structure of the game's formal system, from the social framing of the system as a game (possibly established by the presence of game-like elements that signal a system's "gameness"), or from a combination of the two?

One common explanation as to why games are so engaging and fun to play is that games are highly effective at eliciting a state of flow (Csikszentmihalyi, 1991; Chen, 2007; see also Salen and Zimmerman, 2003; Schell, 2008, among others). Originally introduced by Mihaly Csikszentmihalyi, flow is "a feeling of complete and energized focus in an activity, with a high level of enjoyment and fulfillment" (Chen, 2007). When you are so involved in an activity that you block out the world around you, losing track of time and forgetting to eat or sleep or meet with your dissertation committee—you're in a state of flow. Csikszentmihalyi described a number of key components of flow, many of which are commonly found in good video games, including: clear goals, direct and immediate feedback, and a continuously challenging activity requiring skill. This last component is particularly significant: to achieve flow, an activity needs to be challenging enough that we don't get bored, but not so challenging that we become anxious or stressed (see Figure 3.1). Video games are often particularly good at keeping players in the flow channel by increasing the games difficulty (such as with more difficult levels or opponents) as a player's skill at playing the game grows. Indeed, a number of scholars have explored methods to automatically detect and adapt a game's difficulty in order to maintain this flow state (e.g., Hunicke and Chapman, 2004). Thus when we are playing games, we are attempting an activity that is just the right level of challenge to keep us in flow—and this is what keeps us coming back for more. As Gee says: "Since games are often challenging, but do-able, they are often also pleasantly frustrating, which is a very motivating state for human beings" (Gee, 2003b).

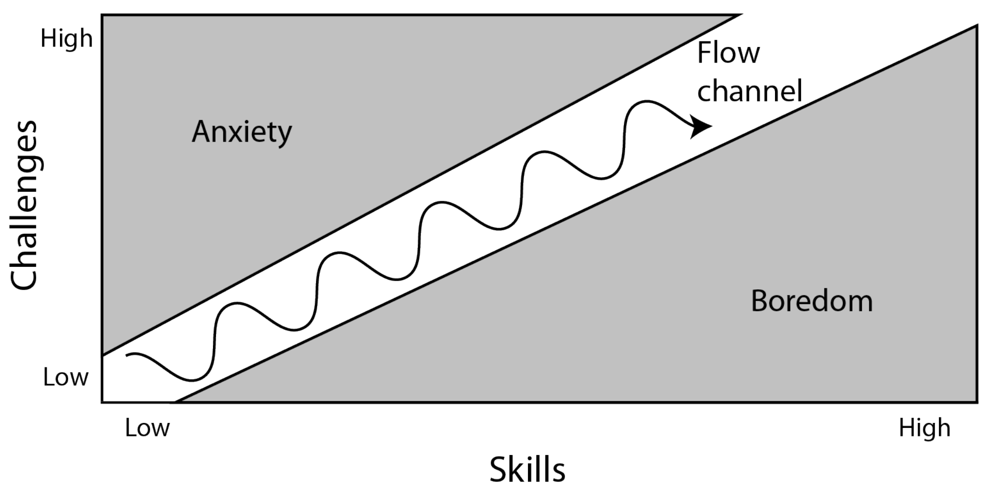


Figure 3.1: The flow channel is the "sweet spot" between an activity that is too easy and one that is too hard (graphic from Schell 2008).

But in order to stay in the "flow channel," we also need to continue developing our skills to match a game's increasing difficulty. Juul claims that "Playing a game is an activity of improving skills in order to overcome these challenges and playing a game is therefore fundamentally a learning experience" (Juul, 2003). When we play a game, we are learning to interact with the rules of the game in order to win—for example, in a video game like *Portal* the player learns to interact with the game's unique simulation of physics and teleportation. And it is this feeling of learning—this sense of mastery and accomplishment—that is the source of the enjoyment and "fun" of being in a state of flow. As Koster says: "Fun from games arises out of mastery. It arises out of comprehension. It is the act of solving puzzles that makes games fun. With games, learning is the drug" (Koster, 2005). Games offer a constant opportunity to learn to "grok" a complex game system and to solve the puzzle of how to play and how to win. Indeed, this is the reason that Schell defines a game as "a problem-solving activity, approached with a playful attitude" (Schell, 2008). Games are non-serious, playful environments for continual learning—learning that is enabled by the constant demonstration and mastery of skills that occurs in a long-term flow state.

Based on my own experience with video games and other activities that produce a state of flow (such as, for me, programming computer software), I find the well-established view that flow is a significant factor in a game's engagement to be highly compelling. By adopting this model, I privilege the level of challenge and the ability to apply skills to a game system as key factors in determining whether a system is engaging or not—rather than, for example, particular design patterns (Björk and Holopainen, 2005) or the dramatic engagement of the narrative. Indeed, an emphasis on flow suggests that a social framing (calling an activity a game) alone does not make a system engaging, but instead the design of the interaction with the underlying system such that the user is consistently and properly challenged. In sum, although I consider whether an activity is a game or not to be socially designated, where a game is engaging or fun to play draws more heavily on a systemic view of games. For this reason, I consider individual system elements and interaction forms as components of a system's game framing in analyzing what characteristics make a game engaging.

3.1.1.1 Evaluating Engagement in Games

There are a number of methods for evaluating whether a designed game (or game-like system) is engaging or not. In part, such a measure correlates with evaluations on whether a game is "good" or not, with the assumption that good games are those that are engaging. Indeed, the strength of a commercial game can in part be measured by its financial success—Gee argues that video games need to incorporate good learning principles (and hence be good games) in order to be successful in a capitalistic marketplace (Gee, 2003a, 2004). The commercial success of games and the knowledge of professional game designers have been used to develop gameplay evaluation heuristics (Desurvire and Wiberg, 2009; Federoff, 2002). Building on strategies found in computer usability studies in the field of HCI, these heuristics include statements of what a good game "should" be. For example, a good game would follow the heuristic "There should be variable difficulty level," matching the principle of providing the right level of challenge and thus enabling flow. A game can then be evaluated by an expert in terms of how well it matches the known heuristic, which itself can be validated against games known to be good (i.e., commercially successful).

In this dissertation, I draw on other game evaluation methodologies that focus more explicitly on measuring the player's engagement with the game system. For example, the GameFlow model (Sweetser and Wyeth, 2005) offers a heuristic based on flow theory (Csikszentmihalyi, 1991)—games are evaluated in terms of how well they support users entering a flow state. The Game Engagement Questionnaire (GEQ, Brockmyer et al. 2009) also considers elements of flow, combined with other measurements of engagement including presence, psychological absorption, and psychological disassociation—all aspects of how involved a player is in the gaming experience. Moreover, the GEQ moves from an evaluation by experts based on a heuristic to a questionnaire methodology—players are surveyed about their agreement with statements measuring engagement (e.g., "while playing time seems to stand still or stop"). The questionnaire method allows for a wider variety of views and opinions about the game's ability to foster engagement to be included, rather than trusting to a small group of experts' application of a heuristic to a potentially theoretical playing experience. Indeed, Fu et al. (2009) use a similar methodology in adapting the GameFlow model to educational games (EGameFlow). I use this form of subjective player evaluation of engagement (with questions based on GameFlow, EGameFlow and GEQ), in evaluating the engagement afforded by the game system presented in this dissertation.

As stated previously, this dissertation considers how game dynamics may be used to support interactive systems for assessing causal understanding; it explores the effects of those dynamics on the causal mapping process used for such assessment. With respect to games, my research questions focus on how game dynamics might affect the assessment procedure and players' engagement with it, rather than "how do we develop a good game for assessing causal understanding?" Developing a highly successful game of any type is exceedingly difficult—particularly when also needing to provide the basis for a dissertation. Following the iterative design philosophy, the game presented in this document is a prototype (and indeed, a relatively early prototype), rather than a final product[6]. Thus while I measure the presented game's engagement levels (based on the research methods described above) in order to consider what characteristics of the game may influence such engagement, the success of the game as an engaging experience is not necessarily a contribution of this research. Game design is hard and has no magic bullet, and so I focus on contributions other than hard and fast rules or techniques for the creation of a successful game.

3.1.2 Serious Games and Gamification

Although many people consider games to be "just" entertainment (indeed, some view video games as a waste of time at best and a cause of anti-social and even violent behavior at worst), games and other interactive experiences are increasingly being studied and used as tools for solving significant social problems—how the engagement of games described above can be used for a variety of purposes. The past decade has seen a variety of research efforts focused on *serious games* (e.g.,

[6]Leonardo da Vinci is often attributed with saying "art is never finished, only abandoned".

Abt, 1970; Sawyer and Smith, 2008) and spin-offs such as "games for change" (as presented at the annual G4C festival, see "Games for Change", nd). Serious games are games that have been "developed with the intention to be more than entertainment" (Ritterfeld et al., 2009)—games that are supposed to be fun while also providing some other value to society, such as in education, health care, or scientific exploration. Games have been linked with non-entertainment purposes for thousands of years (e.g., Halter, 2006), with the 1960s seeing a particular surge in the development of games for education and promoting social cohesion (see Hung, 2011; Pearce et al., 2007). The rise of digital technologies and the popularity of video games over the past decade (Aarseth, 2001) has revived research interests in how to harness games' status as highly-engaging systems. Researchers have explored how to harness the engagement of games for a wide range of purposes—at the most extreme end, researchers such as McGonigal claim that serious games can enable humanity to solve major social challenges such as hunger, economic collapse, and environmental sustainability (McGonigal, 2011). Such claims, while often inspiring, tend to represent a "pie in the sky" view based on the assumption that an ability to master explicitly designed and appropriately scaled challenges (as required for flow) corresponds to the ability to master potentially unsolvable problems that emerge from interactions of vast complex systems in the natural and social world. Nevertheless, it is possible that the engagement of games may be used in more limited contexts, particularly ones where interactive systems can be designed to evoke the required state of flow.

Most recently, an increasing amount of research and commercial interest as been directed at the concept of *gamification* (e.g., Deterding et al., 2011a,b). Deterding et al. (2011a) define gamification as "the use of game design elements in non-game contexts"—that is, adopting elements or characteristics that commonly signify an activity as being game-like. This process most commonly involves the addition of points (and other scoring mechanisms), badges (Antin and Churchill, 2011), or social competition. For example: the location-based social network *Foursquare* gives badges for repeatedly visiting certain places, while the *Nike+* pedometer lets users compete against their friends to see who can run the farthest. In fact, even the fuel-efficiency reporting of hybrid cars such as the Toyota Prius has been described as game-like and a successful instance of gamification. In these cases, the interactive system in question may not be explicitly designated as a game, but include elements that semiotically suggest "this is a game" as a way of making the system more "fun" and therefore engaging. While the concept of gamification can be contrasted with that of serious games (the former refers to using game elements, while the later refers to the design of "full-fledged" game), there is no firm division between the two—the number of game elements needed to distinguish between a gamified system and a full game is highly subjective (see Deterding et al., 2011a, for a more detailed discussion). Thus I use the term "gamification" to refer to the process of applying game framing elements to an interactive system, whether or not the system is explicitly identified by the designer as a game or is simply designated as such by its users-turned-players.

Serious games and other gamification efforts commonly aim to utilize the engagement of games in order to increase participation with a system. In this respect, such efforts are related to the research topic (and parallel term for serious games) of "persuasive games": that is, games that persuade the player to take some type of action[7]. Even game systems designed primary for entertainment typically at the least attempt to persuade the player to keep playing—Ducheneaut et al. describe how *World of Warcraft*'s reward system lets it function as a "virtual skinner box" and condition players to continue playing (Ducheneaut et al., 2006). Persuasive games, on the other hand, attempt to use the motivating capabilities of game-like experiences to motivate some form of behavior charge, in a similar manner to other persuasive technologies (Fogg, 2003). While this change can be as simple as to continue using the game system (as in entertainment-centered games), it often extends to behaviors that fulfill non-entertainment goals. For example, Exertion Games (Mueller et al., 2011) attempt to persuade the player to exercise more by making exercise part of the gameplay, such as by having everyday physical activity power a virtual racer in *NEAT-o-Games* (Fujiki et al., 2008) or a virtual aquarium in *Fish'n'Steps* (Lin et al., 2006). Indeed, many persuasive

[7]This definition is related to but distinct from Bogost's concept of "persuasive games" (Bogost, 2007), which focuses more on games that make a rhetorical argument—often about a social phenomenon—rather than those that prompt specific actions on the part of the player.

games take form as *pervasive games* (Montola et al., 2009), or games that integrate the virtual game world into the everyday "real" world. Adding a game framing can help persuade certain actions by placing them within a magic circle, though the boundary of this circle becomes blurred and requires novel forms of social negotiation. Indeed, a variety of pervasive, persuasive games have been developed in support of environmental sustainability (e.g., Froehlich et al., 2009; Gustafsson et al., 2009; Ross, 2011). Attempts to address sustainability through persuading behavior change has been common in the HCI community, though this approach introduces a number of theoretical and practical problems (see Brynjarsdottir et al., 2012). As such, this dissertation considers instead how gamification techniques maybe used to support (e.g., by persuading participation) assessing understanding of the complex causal systems surrounding sustainability rather than persuading a particular course of action.

Overall, serious games and gamification (as considered in this dissertation) attempt to take advantage of the engagement of games to support non-entertainment purposes. I see engagement and the participation it fosters as the primary benefit of using a game framing. Game dynamics are certainly not the only way to increase engagement—nevertheless, games offer a model (or according to some scholars, a shortcut) for achieving what is a commonly a highly desirable trait of any interactive system or experience: keeping users interacting. Gamification is thus used as a kind of force multiplier: increasing the utilization of a system without necessarily adding new use cases or benefits. Furthermore, gamification attempts often assume that designating a system as a game and making it "fun" (a subjective term that is rarely defined)[8] will increase engagement and thus participation. However, while measures of fun or engagement can be validated by drawing on ideas from flow theory, using engagement as a proxy for participation may miss the requirement to sustain flow over time. A game can be made to be engaging, but participation requires continual engagement—game dynamics that produce engagement in the short term (e.g., during testing or development) may not be sufficiently capable of encouraging participation in the long

[8] Indeed, fun and engaging are not necessarily the same thing (although they are often related in games); Shakespeare's Tragedies are highly engaging, but would not likely be called "fun."

term. Nevertheless, this research considers stand-alone assessments that do not require long-term participation, and thus may be more amenable to gamification techniques that may only produce short-term increases in engagement and participation.

Furthermore, gamification efforts may fail due to the assumption that calling a system a game and giving it elements that signify it as a game will produce engagement. Although such techniques may make a system into a game (following the view that games are systems that are socially designated as such), these elements don't necessarily make it into a *good* game. Indeed (Bogost, 2011) points out that linguistically the "-ify" suffix of the word "gamify" makes the game design process sound easy, straightforward, and applicable to any domain—and it certainly is not any of these. While this argument in some ways may be a reaction to the pervasiveness (even over-use) of the neologism "gamification" in commercial systems, the point stands. Making good games that support a flow state and the resulting engagement is hard, since the engagement quality of a game results more from the systemic properties than from the social framing. Despite years or decades of research, there are few notably successful gamified systems (some which are described in more detail below); there are many more quiet failures than loud successes. The limits of these successes remains an open topic of exploration—thus one of the contributions of this dissertation is a consideration of what characteristics might limit the applicability of gamification techniques in the specific context of educational assessment.

3.2 Games and Education

One context in which serious games are frequently developed and deployed is the domain of education (both formal and informal). Indeed, the use of games—and digital games in particular—for education has an extensive history (see Hung, 2011), producing research that fills multiple long-running journals. As one component of this history, early attempts to use digital games to support education focused on adding a game framing loosely around non-game pedagogical practices. These "edutainment" video games commonly involved "skill-and-drill" procedures and provided a graphical interface for doing math or reading problems (possibly with simple games interspersed); gameplay was thus separate from any pedagogical methods (see Klopfer et al., 2009a, for an overview). Indeed, in these systems the game components could be wholly separate from the educational content—playing the game could simply act as a reward for performing otherwise non-game-based educational tasks. In this way, similar to many gamification projects, edutainment systems would attempt to socially frame an educational activity as a game in order to harness the engagement of games to increase participation. By calling a pedagogical task a game (and giving it the trappings that signify it as such), the task becomes "fun" and thus can encourage participation in practicing educational (e.g., math and reading) skills.

However, more recent considerations of how games may be used to support education has focused in on the potential of game systems themselves to enable learning, rather than just participation in classically educational activities. This view draws in part on the premise that games are engaging because they provide players with the constant opportunity to learn the game system—the fun of games is in fact the fun of learning (Koster, 2005). This understanding of games as systems for learning has led scholars such as Gee to refer to them as "learning engines"—systems for teaching players how to play (Gee, 2004). In his seminal book, *What Video Games Have to Teach Us About Learning and Literacy* (2003a), Gee identifies a number of sound learning principles that are incorporated into the characteristics of good video games. These include the principle of working at the edge of the flow channel so tasks remain "challenging but do-able", but also principles such as providing "just-in-time" information, creating an environment for active experimentation, and allowing the player to take on the role of an expert at solving the game's problems. In this view, games are systems that excel at enabling player learning of that system (though a combination of both systemic structure and social framing); thus if the system or skills to be taught can be directly mapped to the game system, then in learning the game the students could also learn the desired content. This argument posits that games provide a context and framing that has been demonstrated to be highly effective at supporting learning—the challenge becomes how to make sure the content being learned is that desired by a particular educational context.

Moreover, games' characteristic of being effective learning engines is further enabled by games' position as simulation systems (see e.g., Squire, 2003). Digital games, in addition to being interactive systems, also commonly reflect a simulation of some sort: as formal systems, games establish a model of how the rules and the "world" of the game work. In this way, game frames create a concrete context in which interactions with the system occur. This context is most notable in a game's narrative framing (e.g., Jenkins, 2003), but can also be understood simply as the context of the complex system that underlies the game's interaction. This simulation provides a context for the learning that occurs within games (i.e., the context of learning the system underlying the simulation)—a context that enables forms of concrete, situated learning (Shaffer et al., 2005; also Lave and Wenger, 1991) that can better support the use and transfer of learned knowledge. Shaffer relates this context to the idea of an "epistemic frame" (Shaffer, 2006)—a frame for organizing understanding within a particular community of practice. Game systems can support learning by grounding it within the concrete context of that system, though the details about whether and how this contextualized understanding can effectively transfer outside of the game context is a topic of ongoing research. Note also that the effectiveness of games' contextualization also enables players to play at assuming the identity (see Ito, 2010) of a member of a particular community of practice[9]. This temporary alteration of the presentation of self (Goffman, 1959) can support increases in players' sense of self-efficacy (Bandura, 1997), which can have beneficial impacts on learning.

By considering games as systems that situate learning with a concrete context, games can also be seen as supporting education by more effectively enabling *active learning* (Bonwell and Eison, 1991). As interactive systems, games rely on player actions—players need to interact with the system in order to learn its underlying rules structure and thus to perform situated learning. Pedagogies in which students take an active role have been repeatedly demonstrated to be more

[9] Players can even assume fantastical identities; McGonigal (2011) argues that the way in which games allow players to regularly perform the accomplishment of "saving the world" supports individual empowerment that can encourage both learning and social action.

effective (e.g., Prince, 2004) at supporting learning by helping students in constructing their own understanding of a topic (Piaget, 1930; Papert, 1980). As one example, *Gamestar Mechanic* uses a game framing to create an environment in which players develop their own games, thereby learning the process of game design (Salen, 2007). In this way, by acting as interactive systems that contextualize framing for learning, games provide an effective scaffolding for enabling forms of active learning—namely, the learning that actively occurs through playing the game. Games support education by providing systems and contexts for active learning.

Although games as learning engines can support education by situating learning within concrete contexts, there are still other recent efforts that continue to draw upon games as tools for increasing engagement (Squire, 2003) and thereby participation in educational activities. For example, groups such as the Digital Media and Learning Initiative have considered providing badges as a way to inspire learning (DML, 2011); in this case, a game-like element (badges) are being used to signify that education is a game, thereby drawing upon similar arguments as made by gamification proponents that if an activity looks like a game, it may be engaging. Such increased engagement is often offered as supporting learning by leading to an increase in effort and attention on the part of the student. As Squire and Jenkins suggest: "The properties and processes of a well-designed game may motivate [students] to turn to textbooks with the intention of understanding rather than memorizing" (Squire and Jenkins, 2004). Thus efforts to use games to simply increase engagement with educational material may rely on the assumption that further effort can best support learning, rather than perhaps reframing the material or skill being learned or other fundamental approaches to pedagogy. Along similar lines, professors have experimented with treating college classrooms as games themselves (e.g., Travis, 2009; Sheldon, 2011), and even entire schools have been redesigned as gaming platforms (see Quest to Learn, 2011). Nevertheless, these gamification efforts continue to draw on the social framing view of games (making an activity a game through social designation, rather than the design of the underlying system), experimenting with utilizing the engagement of games for education to drive participation.

Overall, previous researchers have positioned game dynamics as being often highly appropriate for supporting some forms of education. This effectiveness arises from two sources: games are able to increase engagement and therefore effort (similar to arguments for gamification in general), and that games as learning engines provide contexts in which situated learning can occur and in which students can more readily practice and demonstrate learned skills—games better enable students to "learn by doing." Although there is a long history of research efforts exploring these potential benefits, such research is still ongoing: scholars are still exploring how to best harness these benefits in games, and how to evaluate the effectiveness of such benefits. Other issues, such as role of the educator in game-based education, also are still under consideration. Nevertheless, this dissertation takes as a launching point the assumption that games may be capable of supporting education—even if the details on how best to effectively harness that support may be under development.

3.2.1 Games and Systems Education

Of particular interest in this dissertation is previous work in using games to support *systems education*—that is, teaching and training students in the process of systems thinking. As described above, the multiple interactions between rules and potential moves means that even the simplest games are themselves complex emergent systems. For example, the ancient game of Go has only a few written rules—players take turn placing stones on a 19x19 grid, and a stone surrounded by your opponent's is lost—yet elicits such complex emergent strategies and gameplay that computers have only recently been able to compete at the professional level (Lee et al., 2009). In this way, games represent complex systems that players interact with and learn to master. Thus the underlying structure of a game offers a suitable basis for modeling other complex systems to educate students about; in learning to understand the complex game system, students can learn to understand the complex system that the game maps to and represents. Games may represent a suitable venue for supporting player exploration and processing of complex systems (such as through cognitive mapping), as players are already engaged in systems thinking simply by considering the rules in order to develop strategies to play (see Salen, 2007; Squire, 2002). Indeed, previous games have even represented complex systems of causal relationships: the classic board game *Mouse Trap* allows players to toy with the idea of causal paths in constructing the titular, Rube Goldberg-esque trap. Thus overall, viewing games as learning engines suggests that their underlying complex systems have enabled them to serve as an effective basis for enabling active learning about and exploration of causal systems.

As mentioned previously, one prominent and early example of using a game for systems thinking education is the MIT Beer Game, created at the MIT Sloan School of Management in the 1960s (see Goodwin and Franklin, 1994; Sterman, 1989). I provide here a detailed description of this game to demonstrate how the rules that define a game's complex system can demonstrate systems thinking concepts, as well as providing an example of a successful game for education in this domain. In the MIT Beer Game, players take on the roles of different actors in a beer production and distribution business. Players are assigned to different positions in this business: factory, distributor, wholesaler, and retailer. Customer demand drives the game—the retailer is told this demand and orders beer from the wholesaler, who orders beer from the distributor, who orders from the factory, who can produce unlimited amounts of beer. Players only know what orders are coming in; they are purposefully not given a complete view of the system. Moreover, a key component of the game is that orders and shipments each require two time periods to reach their destinations, introducing delays into the process. Each player has a single goal: minimize all costs, which are incurred either by holding inventory ($0.50/case/period) or by back-ordering ($1.00/case/period)—thus there is incentive to carry extra inventory rather than back-order.

The game begins with the system balanced: every player has an inventory of 12 cases, and there are 4 cases in transit between each player. The customer demand starts at 4 cases and remains steady for 4 periods, keeping the system stable. Then in the fifth period, customer demand increases to 8 cases and remains there for the rest of the game. This single small change introduces drastic changes to the system's balance:

> "[the change] is sufficient to generate a pattern of increasing demand that cannot be met, depleted inventory levels, rapidly increasing back-orders followed by deliveries of large quantities of beer. As the large shipments arrive, demand suddenly declines and inventory levels increase rapidly... It is very common to see individual orders, at all of the positions, for as much as 50 cases of beer per period, and both inventory and back-order levels reaching the 50 to 75 case range. On occasion, orders for a single period have climbed as high as 500 cases, with inventory and back-orders rising to similar levels" (Goodwin and Franklin, 1994).

Furthermore, as Goodwin and Franklin report: "the Beer Game has *never* failed to work" (original emphasis). This simple change produces severe, ripple effects throughout the causal system that the players do not expect. Indeed, this effect is so unexpected that players often initially conclude that the game is "rigged." The usual purpose of the game is to "shock" players into realizing that local causes (e.g., how much beer they order) can have significant global effects, simply due to lack of information and a failure to recognize the game's causal feedback loop. The Beer Game thus demonstrates one way in which games are used to teach systems thinking: to reveal common misunderstandings of the interactions within complex and dynamic systems. By enabling players to directly interact with (and indeed, become a part of) the complex system, they can develop a firmer understanding of the systems concepts that lead to particular behaviors—in this case, the effects of delays on a complex system. In this case, the game framing introduces additional rules (e.g., the to-win conditions) that ensure the desired pedagogical behavior.

While the Beer Game reveals misunderstandings of a complex system, other game-like systems have previously been developed that help users to construct and experiment with their own models of complex causal systems. One of the most commonly used software tools is STELLA (Richmond, 2001) developed by isee systems. STELLA enables users to construct dynamic causal models of complex systems, give relationships and factors different weights or values, and then experiment with how these changes affect the system. Users are often encouraged to "play" with the system, treating it as a kind of simulation game potentially similar to commercial games such as *SimCity*. As an example using content near and dear to my heart, educators have used STELLA to model character relationships in Shakespeare's Hamlet—students develop a model of the causal factors that lead to Hamlet refusing but eventually giving in to killing Claudius (Hopkins, 1992). Students include as factors various points of evidence that Claudius killed King Hamlet, how that evidence increases the factor of the prince's motivation, and how that motivation eventually and dramatically decreases the factor of Claudius' lifeline. In the process of constructing this model, students discussed the components of the play and the characters' motivations, improving their understanding of the literature. Thus by using the software tool to reframe the discussion of literature in terms of systems thinking, the students not only gained a deeper understanding of the complex social system present in the Bard's work, but also learned to apply the lens of systems thinking in a novel medium. This example demonstrates how simulations and games built around complex systems can be used for education, beginning to validate the view of games' effectiveness discussed previously.

Relatedly, other research has considered how students can experiment with constructing models of complex systems through the process of constructing the complex system of a computer game itself. For example, StarLogo TNG (e.g., Begel and Klopfer, 2007; Klopfer et al., 2009b) is a programming environment for developing games using agent-based programming. Based on the StarLogo system (which was used for visually modeling complex, dynamic systems), StarLogo TNG lets students model emergent complex systems by defining the behaviors of the system's agents, and then creating a game out of interactions of those agents. By allowing students directly to create and interact with the system in the process of building the game, they gain a greater understanding of the system dynamics that lead to such emergent behavior. Similarly, *Gamestar Mechanic* (see Salen, 2007; Torres, 2009) uses a game framework to teach people the process of game design—that is, how to build the complex systems that are games. In these cases, students still interact with the underlying game system—with the same benefits—but do so from the standpoint of a designer rather than of a player. These examples further demonstrate how as complex systems, games offer an effective basis for teaching about forms of systems thinking (in this case, the systems thinking that embedded in game design and programming).

All of these examples draw upon how games (and particularly video games) are interactive simulations. Sterman (1994) calls computer simulations "essential" for testing causal maps, and describes some of the advantages of virtual worlds for effective learning—including trying potentially dangerous experiments (e.g., when learning about the complexities of flight) and providing high-quality feedback. Following a constructivist view (Papert, 1980), this active experimentation and testing is what supports developing system thinking skills for understanding complex causal systems. Nevertheless, a key component of this active learning process is the *construction* of the system model—students cannot just run tests on existing models, they need to experiment with building the models themselves (Goldstone and Wilensky, 2008). As games prioritize player interaction, they can be seen as particularly suitable for enabling this form of active learning. And because systems thinking can be effectively linked with the complex system underlying a game's structure, systems education may be an especially suitable domain for games and simulations. However, it is possible that framing a simulation system as a game (with the assumptions of play and the separateness of the magic circle implied by such a framing) may influence how players interact with the simulation. In this dissertation, I consider how games as a socially designated form of interaction may influence how a player interacts with the game's underlying simulation of a complex system—particularly in the domain of educational assessment.

3.2.2 Games and Assessment

As described above, games have been positioned as able to effectively support education. But what about educational *assessment*? Assessment is a key component in many pedagogical intervention, both for identifying what to teach (i.e., establishing learning objectives) and for determining if students actually learned what was taught. If assessment is a component of education and games are effective at supporting education, then it follows that games may be effective at supporting assessment. It is possible that the increased engagement found in educational games may be able to encourage further participation and effort applied to the assessment process, thereby leading to more accurate measures. The situated context provided by a game may also allow for assessments to be performed with more demonstrable measures (i.e., skills applied to a particular situation). Moreover, the non-serious magic circle established by a game may be able to make an assessment more "fun," thereby reducing text anxiety (Hembree, 1988) and similar confounding barriers. Games have the potential to support assessment just as they may support education more broadly.

As games are frequently computerized systems, performing assessment using video games can be seen as a form of computer-based assessment. Computer-based assessment has been the subject of a significant amount of research (see Thelwall, 2000, for one review). Assessments that are partially automated via computer systems can be prominently seen in the form of standardized tests (e.g, using Scantrons) found throughout formal education institutions—such computer-based assessments are able to support the automatic graded of multiple-choice exams. Moreover, Thelwall describes how computer-based assessment can support formative open access tests, in which students *self-assess* their own learning. In these ways, computer systems are primarily used to support the *logistics* of assessment, either by speeding up grading or by providing greater access, rather than necessarily improving the assessment process itself (though automating steps such as question randomization may be a start). Indeed, there are concerns about whether computer-based assessment can effectively measure the higher learning principles (i.e., higher stages in Bloom's taxonomy; see Bloom et al., 1956) normally assessed through more qualitative tasks such as essay writing. While the topic of automatic essay scoring is an area of ongoing research (e.g., Shermis and Burstein, 2003), the capabilities of computer-based assessments to support the subjective nature of many educational assessments have not been established. Note that in this dissertation, I focus on assessing the identification and knowledge[10] of causal relationships—lower-stage learning principles that are likely more amenable to computer-based assessment techniques.

[10]Though the question of how to assess knowledge remains under debate (e.g., Anderson et al., 2000; Bloom et al., 1956).

Although computer-based assessments are well-established, the use of computer games for assessment is currently relatively underexplored (though see Shute et al., 2011). Research in assessment and educational games has focused on considering how to integrate assessment processes into learning games, performing embedded or "stealth" assessment (Shute, 2011). Stealth assessment in games means that the assessment occurs without the player being aware of it—players interact with the learning game as normal, while data that can be used for assessment are collected in the background. Such assessments thus draw upon measurable elements of the game, following an evidence-centered approach to assessment design (Shute et al., 2009). Stealth assessments are intended to maintain flow in learning games, thereby not interfering with the engagement and learning enabled by the constant challenging that occurs in a flow state. Moreover, such embedded assessments can be used to provide the constant feedback necessary to maintain a flow state and make a game "good"—indeed, much of a game's feedback (such as how much health an avatar has) can be seen as a kind of embedded assessment of a player's skills in that game. Applying stealth assessment to education theoretically just involves altering the parameters used to measure and provide feedback (e.g., reporting accuracy of systems thinking rather than amount of health left) (Shute, 2011).

Stealth-based assessments are most commonly applied to immersive learning games, such as virtual-world like simulations of a system or event. As a typical example, Sliney and Murphy (2011) describes a simulation game in which the player takes on the role of a junior doctor who needs to diagnose patient conditions while "on call" at a hospital. The game nature of the system provides the player with a low-risk environment in which to experiment with and practice at being a doctor, while assessing the player's skills at performing this role through the analysis of actions (moves) taken within the game. Note that with this form of embedded assessment (and indeed in any assessment), one of the key concerns is determining what skills need to be measured and how to measure them. For this reason, other research regarding assessment in the domain of games for education focuses on the assessment *of* game-based learning—that is, how to measure the skills that are being taught in gamified learning systems (e.g., the gaming and 21st-century literacies targeted in *GameStar Mechanic*; see Salen, 2007, Games, 2008), regardless of the technique used.

Overall, previous research has considered the relationship between games and assessment in the context of integrating assessment into existing learning games. However, there has not yet been a consideration of how games may be used to support *stand-alone* assessments—assessment systems designed only to test rather than to also teach. Stand-alone assessment systems may be relatively easier to adopt in a variety of contexts: stand-alone assessments may be more easily integrated into existing formal education curricula, and the shorter interaction period required for assessment may support adoption and use in non-educational contexts. Moreover, studying stand-alone assessments can be an important step in exploring the relationship between assessment and games, apart from any confounding influences of a game's teaching-focused components. As such, this dissertation focuses on the novel consideration and development of a stand-alone interactive system and game for assessment, measuring the lower-level yet abstract skill of identifying causal relationships (rather than more concrete skills that may be detectable in more immersive games). Nevertheless, I draw on some of the concepts of stealth assessment in socially framing an interactive system as a game rather than a test, in order to support the described benefits of games and their magic circles for assessment.

3.3 Gamification Approaches

Previous work has demonstrated a variety of approaches to gamifying systems (see e.g., Deterding, 2011; Sawyer and Smith, 2008), ranging from the inclusion of particular game-denoting elements (e.g., points or badges) to specific forms of interaction or media. In this section, I discuss some of the approaches used to develop gamified systems (primarily outside of the domain of education), as illustrated through specific examples, flagging approaches for developing games that may be suitable for supporting the assessment of causal understanding.

3.3.1 Human Computation Games

One form of gamified system that I see as particularly relevant to assessment is that of *human computation games*, or *games with a purpose* (GWAPs) (von Ahn, 2006; von Ahn and Dabbish, 2008). These games are serious games "in which people, as a side effect of playing, perform tasks computers are unable to perform" (von Ahn and Dabbish, 2008). GWAPs thus draw upon (and even helped to found) the field of "human computation", in which humans solve cognitive tasks that cannot feasibly be completed by computers (see Quinn and Bederson 2011 for a recent overview of this field). For example, human computation is commonly applied to the domain of computer vision for tasks such as image labeling. Determining the semantic content of an image is generally incredibly fast and easy for humans (e.g., answering: "Is this a picture of a cat or a dog?" "Is this a picture of a car or a tank?"), but remains a difficult and unsolved problem for computers to address algorithmically. Thus human computation is used to distribute such problems to humans for solving. GWAPs gamify this technique, using gameplay to motivate humans to complete these often-small and otherwise monotonous computational tasks—again, drawing on the assumption that "engagement leads to participation."

The foundational GWAP is the *ESP Game* (von Ahn and Dabbish, 2004)—a game that effectively motivated players to label a vast number of random images through the simple fantasy narrative framing of trying to read someone's mind combined with a scoring system to determine how well the players did. Since then, GWAPs have been developed for a variety of purposes. As one example, *Plummings* (Terry et al., 2009) uses a narrative framing (of moving cogs and pipes to supply oxygen to a colony of creatures) to have humans attempt to efficiently place Field Programmable Gate Arrays (FPGAs) and minimize logic path distances. *PhotoCity* (Tuite et al., 2010) combines a GWAP with a pervasive augmented reality game—players play a game of "capture the flag" by taking pictures of buildings that best generate virtual 3D models of that building, thereby supporting the digital reconstruction computer vision problem. One of the most successful serious games is a human computation game called *Foldit*, a puzzle game where the player bends, shakes, and otherwise manipulates protein structures to try and minimize the energy in the protein's shape—in effect to perform protein folding. By playing this game Foldit players have helped discover protein models that could aid in the fight against HIV and other retroviruses (Khatib et al., 2011b). Indeed, the game players are even helping to develop new algorithms for protein folding directly based on their game strategies (Khatib et al., 2011a).

These examples illustrate two basic forms of human computation games. Games such as the *ESP Game* and *PhotoCity* involve humans providing information that computers are unable to generate algorithmically—the *ESP Game* collects descriptions and labels of images, while *PhotoCity* collects pictures of buildings and control points for reconstructions. On the other hand, *Plummings* and *Foldit* have humans apply their own mental heuristics to optimize existing models (of FPGAs or proteins, respectively); because humans may be able to identify patterns or make intuitive leaps that machines cannot, people may be better at these manipulation tasks than computers. While the second of these forms (particularly with *Foldit* as an example) does demonstrate the possible success of GWAPs, I believe that the first form may offer a particularly suitable structure for developing games for assessment. A game for assessment could collect information from players—specifically information that can be used to measure their understanding. I adopt such a design in this dissertation, presenting a game in which humans perform the task of identifying causal relationships (similar to human computation tasks that might focus on natural language processing), but rather than using these data as input to a computational system, I use them to evaluate the player's understanding of that causal system.

As an aside, it is worth noting that the concept of human computation has a long history leading towards its use through games. In fact, human computation used to be the only form of computation: "computer" was a job title given to people (often women) who performed mathematical calculations, computing distributed sums and products in order to solve complex problems such as interplanetary physics, ballistics, or to otherwise further scientific knowledge (see Grier, 2005; Light, 1999). For example, the Mathematical Tables Project (founded during the 1930s as part of government relief efforts for the Great Depression in the U.S.) had hundreds of workers computing tables of logarithms, probabilities, and other mathematical facts to be used by statisticians, surveyors, engineers, and others (Grier, 2005). Although machines and digital systems soon replaced these professional human computers, in recent years there has been a revival of systems that use human thinking to perform computation. Benkler (2002) argues that the ease and breadth of modern communication technologies (e.g., the Internet) has enabled large groups of people to easily organize and choose activities (such as playing a game with a purpose) to perform, in what he calls "distributed peer production." Thus it is now possible to develop games and other systems that involve large numbers of human computers—and thus potentially to develop systems that can assess understanding of a large population.

There are methods other than games that can be used for motivating humans to solve computational tasks. Some human computation methods rely on contributors to develop their own intrinsic reasons for working without the framing of a game—people choose to contribute to Wikipedia because of personal motivates, such as wishing to contribute to something larger than themselves (Bryant et al., 2005; Kuznetsov, 2006). Indeed, the same intrinsic motivations that support Wikipedia, such as an increased sense of autonomy or of having contributed to a whole, can also exist in GWAPs such as *Foldit*. However, the most common method of gathering human computers is to forgo the intrinsic rewards of games and instead offer extrinsic rewards—that is, pay large groups of people to work on a problem through crowdsourcing (Howe, 2008) systems. Indeed, crowdsourcing techniques have been applied to aspects of assessment: Heimerl et al. (2012) describe a system that successfully crowdsourced grading of exam problems. Moreover, there are a number of micro-task marketplaces such as Amazon Mechanical Turk[11] (Kittur et al., 2008; Ross et al., 2010b), or MTurk, in which humans can complete small but computationally difficult tasks (such as image identification or natural language processing) for small amounts of money—MTurk pays

[11] `http://www.mturk.com`. Mechanical Turk is named for an 18th century automaton invented by Wolfgang von Kempelen. This machine traveled through Europe, stunning audiences for 50 years with feats of mechanically calculated chess playing prowess. However, the automaton was a hoax—a hidden human controlled the chess-playing machine, thereby faking what was believed to be a thinking machine. Thus the tagline for Amazon's system is "artificial artificial intelligence."

as little as $0.01 for a quick task up to a few dollars for more involved jobs, such as transcribing audio clips. Despite the relatively small amount of pay, such systems are highly successful and are actively used both for solving human computation problems and performing a variety of research studies (Kittur et al., 2008); indeed, the evaluation presented in this dissertation is performed in part through MTurk.

A further benefit of the distributed human computation found in crowdsourcing systems and GWAPs is that such problems can take advantage of how people process information in unique, yet correct, ways: alternative interpretations of information (like labels for an image) may lead to a fuller and better solution to the problem. A large group of diverse people acting independently can potentially produce more intelligent results than any individual member, in an effect popularly referred to as the Wisdom of Crowds (Surowiecki, 2005). This kind of crowd wisdom enables people to collectively process information in collaborative filtering systems (e.g., Goldberg et al., 1992; Ross et al., 2010c) or the creation of systems such as Wikipedia. In fact, this kind of collective intelligence may be particularly effective when gathered through the collaborative framework of a social game (see McGonigal, 2007). The research presented in this dissertation aims to lay the groundwork for collectively gathering and measuring causal knowledge regarding environmental systems in order to potentially improve the modeling and collective understanding of these complex systems (see Chapter 8. Future Work).

3.3.2 Other Gamification Approaches

In addition, here are numerous other forms of gamification other than human computation games. Many serious games take the form of simulations, allowing the player to play at a certain role or performing a certain action within a virtual world. For example, *America's Army* is a hyper-realistic first-person shooter in the vein of *Modern Warfare* or *Tom Clancy's Splinter Cell* that draws on actual military equipment and tactics used by its developer, the United States Army.

Originally released as a recruitment tool (Halter, 2006), the game has since formed the basis for in-use training simulators—indeed, there is a long history of using games for military training (see e.g., Garris et al., 2002; Shaffer et al., 2005). Other simulation-based serious games draw on the way that simulations enable players to interact with the underying complex system. From the domain of environmental sustainability, *Super Energy Apocalypse* (Doucet and Srinivasan, 2010) involves the players developing a sustainable energy production strategy as they try to fend off a post-apocalyptic zombie invasion[12]; because pollution strengthens the zombies, players need to figure out how to most efficiently power their defenses while reducing the amount of smog generated. Indeed, there are a number of simulation games in which players practice energy and pollution management, even without the apocalyptic narrative—research efforts such as *SCAPE* (Podleschny, 2008), teaching tools such as *ElectroCity* and *EnergyVille*, and even commercial efforts such as the venerable *SimCity* series. These games' performance frames allow players to experiment with different strategies in a risk-free context (that is, they will not actually destroy a city by making bad choices), thereby potentially supporting their autonomy and learning (Gee, 2003a).

All of the above simulation games are perceived as games (rather than just computer simulations) in part because of their game-like narrative framings. Indeed, narrative is one of the most important signifiers that an activity is a game. Many game narratives adopt fantastical elements (e.g., mind-reading or zombies) that help set the game apart from the "real" world and thereby establish a magic circle. By framing a system with non-realistic elements, that system is in a way designated as playful and thus game-like—and thereby encourages a playful approach to this system that is socially designated as a game. Gamification thus often relies heavily on the use of game narratives to reframe everyday, non-game activities as playful. For example, *Chore-Wars* uses a game narrative to let the player take on the role of a fantasy adventurer and earn Experience Points for doing the dishes. In this case, the game's framing narrative is used to turn what is normally "work"

[12]Indeed, other research (e.g., Tomlinson et al., 2012) is exploring how technologies and interactive systems for supporting sustainability in the context of imminent societal collapse (though zombies are not likely to be a concern).

into "play." The use of game narratives is a significant approach for gamification because of how they can encourage the creation of the magic circle and a playful approach to the game. Moreover, game narratives can be used to mask the assessment nature of an activity, thereby support stealth assessments. Adding a game narrative can be seen as a designer explicitly designating an activity as a game, thus beginning the social process of making an experience into a game. As such, this dissertation makes heavy use of game narratives in its exploration of gamifying assessment.

Narrative also plays a large role in the *Karunatree* project (Lyons et al., 2012; Lyons and Tomlinson, 2010; Ross and Tomlinson, 2010)—a system from which this dissertation draws significant inspiration and grounding. *Karunatree* is composed of a narrative-centered game, web application, and pedagogical curriculum designed to support children's understanding of the webs of cause and effect that exist between human activities and natural ecosystems. The game's narrative tracing of the causal relationships between the everyday actions of children in Orange County and the ecosystem of the Sumatran rainforest has provided a foundation for my own consideration of causality, and how maps of causal relationships can be created and understood. Moreover, the use of narrative in this project helps to indicate the relationship between narrative (linear or otherwise) and how people consider and model the causal relationships of complex systems.

Finally, as described above, serious games may use such narratives and other gamification approaches in attempting to persuade behavior change. With this approach, the game may simply have a narrative or interaction style designed to persuade the player to adopt a particular point of view (and thus behavior). For example, *Re-Mission* is a third-person action game in which the player moves an avatar through a 3D environment representing the human body, trying to locate and shoot cancer cells with weapons based on common treatments such as chemotherapy. Nevertheless, research shows young adults with cancer who play this game are more likely to follow with their treatment schedules, as well as gain higher levels of confidence and knowledge (Kato et al., 2008). Other games, particularly in the domain of environmental sustainability, may pull from self-tracking and eco-feedback technologies (see Froehlich et al., 2010; He et al., 2010; Li et al., 2010), converting collected "real-world" data into inputs to the gaming context. For example, in *EcoIsland* (Shiraishi et al., 2009) a family of players try to keep their virtual island home from flooding by reducing their actual self-reported CO_2 emissions, while in *UbiGreen* (Froehlich et al., 2009) players try and raise a family of virtual polar bears by using more environmentally-friendly modes of transportation[13]. Similarly, *EcoPath* (Ross et al., 2010a) used the locations of recorded sustainable behaviors as inputs to a game where players tried to conquer territory in their local community. On the other hand, *Power Agent* (Gustafsson et al., 2009) mixes these two approaches, combining a classic platformer video game with a pervasive system—players play the platformer to earn clues about how they can reduce their home energy usage in competition with other families. These games are focused on directly encouraging sustainable behaviors, rather than simply simulating or rehearsing behavior change (Stokes et al., 2010). While these examples demonstrate how gamification has been applied to the domain of sustainability, I see such approaches to persuasion to be less relevant for stand-alone assessments in which the goal is to gather information about player understanding rather than persuade them to a specific mode of action.

This section has included just a small sample of previously developed serious games and gamification approaches. Indeed, a wide variety of game patterns (Björk and Holopainen, 2005) and genres (Apperley, 2006)—from shooters to puzzles to roleplaying games—have been used in serious games and gamification to achieve societal goals beyond entertainment. But for the purposes of assessing causal understanding, I believe that the structure of a human computation game with the system's game-ness established primarily through a narrative framing may be the most suitable approach to developing a prototype interaction system in the form of a serious game.

[13]Note that although its designers did not frame *UbiGreen* as a game, many users described the system as being a game.

3.4 Conclusion

In this chapter I have provided an introduction to the concepts of games and play, as well as to the fields of serious games and gamification. I detailed how and why games have been positioned as highly suitable for supporting education in general and systems education in particular, yet pointed out that there has been relatively little research into using games for assessment as a component of education. Therefore in this dissertation I offer an exploration of how elements of gamification may fit with interactive systems for performing stand-alone assessments.

In this dissertation I draw upon two broad but separate topics—causality and game design—and combine them to explore how interactive systems and games may support assessing causal understanding. This dissertation thus can contribute to the fields of causality and systems thinking by applying an interactive system and game design perspective, and can contribute to the fields of serious games and game design by considering the appropriateness of particular game elements for the specific but sizable domain of assessment. Thus my research takes two disparate fields, and uses each to inform the other in the development of a new computational system—a prototype of which I present in the next chapter.

Chapter 4. *Causlings* Implementation

In this dissertation, I present a prototype of an interactive computer system for assessing users' understanding of causal relationships in complex systems, applied to the domain of environmental sustainability. The system takes the form of a digital serious game called *Causlings*. This game is based on the technique of cognitive causal mapping, applying the CMAST methodology used by Plate (2006, 2010) through the medium of an interactive computer game—this project is an automated and gamified version of Plate's assessment. With this research, I explore the potential to automate the causal mapping process and thereby gather additional data on the process by which users construct cognitive causal maps of complex systems in order to better measure causal understanding. In addition, by framing this interactive system as a game, I begin an exploration of how game-based systems may be able to support assessment as well as education more broadly (as in Gee, 2003a). Overall, this research seeks to answer the questions of what interactive systems can reveal about the cognitive mapping process, and how a game framing may influence assessment of causal understanding.

This chapter gives an in-depth description of the prototype *Causlings* game system. I begin by describing an overview of the system from a player's perspective, explaining how the game is played and how this play builds on and automates existing forms of assessment. I then give a more detailed consideration of the system's implementation as a game and the design of a number of key components.

4.1 The *Causlings* Game System

Causlings is a web-based interactive computer system[1] in which users (hereafter: players) construct causal maps by linking causal factors with causal relationships, following the structure of the CMAST assessment technique. At the same time, *Causlings* frames the CMAST technique as a game form of interactive computer system with the addition of a game-like narrative. Rather than simply having subjects attempt to choose the direction of causal relationships when specifying a causal map, the game casts this process as one of building bridges (representing relationships) between islands (representing factors) so that a group of virtual creatures called Causlings can visit new islands. The Causlings are only able to cross bridges that correspond to accurate causal relationships (in comparison to expert-generated maps)—thus players need to use their understanding of the causal relationships in the complex system in order to help the Causlings and succeed at the game.

Upon first starting the game, the system presents this narrative framing to the player in the form of a text-based story:

> *The whole population of Causlings is in dire need of your help!*
>
> *"What exactly is a Causling?" I hear you ask? Well they're tiny island-dwelling creatures who dream of visiting every island in the sea! And as no Causling ever learned how to swim, they need you to build them bridges from island to island.*
>
> *The Causlings begin at the fabled Causling Capital, and will happily try to cross any bridge you make for them. But each bridge can only support a few crossers at a time, so you'll want to make as many bridges as you can. When a Causling reaches a new island, it will build a new Causling settlement and quickly raise more young Causlings. Predictably, these young'uns will also want to start exploring, and so will soon leave home to visit other islands across the bridges you've made. But you'll need to act fast: Causlings can be impatient, and may try and swim if there aren't any bridges open for them to cross (and we all know that swimming does not end well for them!)*
>
> *Sound easy enough? Well of course there's a catch. Problem is, Causlings are impeccably logical creatures, and so they can only travel across causal bridges—otherwise*

[1]The *Causlings* system prototype is currently available at `http://thiscausesthat.org/game/`.

they are hit with overwhelming vertigo and tend to fall into the sea. "Causal bridges?" you ask with growing trepidation? See, each island in the sea has a strong controlling concept, and the only bridges that are stable enough for the Causlings to cross are those that match a cause-and-effect relationship between the islands' concepts.

The joking tone of this story aims to establish the playful nature of the game, emphasizing that the activity the player is about to perform is a game and not work (even though, as *Causlings* is a form of GWAP, the player is in fact doing cognitive work to generate the causal map which the system can automatically assess). Furthermore, this story places an initial emphasis on the creatures and their problems, instead of on the idea of a causal bridge and the mapping process, in order to help create an empathic relationship with the virtual creatures. Such empathy can allow the game narrative to provide intrinsic motivation, as players may be more willing to play because of an altruistic desire to help someone in need (Ryan and Deci, 2000)—even simple game narratives often involve helping an avatar to achieve a goal. Indeed, the text description is accompanied by images of the Causlings as adorable humanoid islanders (see Figure 4.1), which can help to further establish this empathy and motivation.

Figure 4.1: An illustration of the Causlings (by Alicia Chan).

After this introduction to the prototype, the player is given a longer text-and-image-based explanation of how to specify causal bridges—that is, how to include causal relationships in a causal map. This explanation is also adapted from Plate (2006), and uses a small example with four causal factors to describe how to define bridges with the appropriate causal direction (see Figure 4.2. This aims to act as a kind of "manual" for the game; the mouse-based interface for creating a causal map through the interactive system is presented pictorially later, just before play begins.

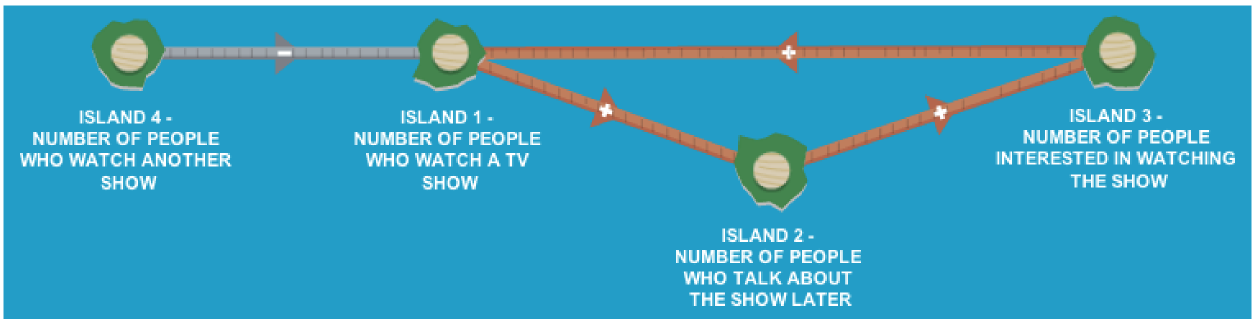


Figure 4.2: Players are shown an example of using increasing relationships ("positive bridges"), decreasing relationships ("negative bridges"), and causal loops–the last of which is not explicitly called such.

For the prototype system, the player is then given a short (1000 word) article describing a controversy surrounding overfishing of a fictitious species of fish called the samaki. This article provides a basis of knowledge for the player to use in building causal bridges (and for the game to assess—see below for discussion of this process). Note that this series of instructions and articles does introduce a steep learning curve to playing the game, as well as requiring a significant amount of reading before the player can actually being to play. However, this procedure mirrors Plate's CMAST methodology, allowing the *Causlings* game to make use of the assessment validity established by this previous research. By directly building on prior techniques for assessing causal understanding, my system is able to directly apply automated assessment and to explore the addition of game mechanics with less chance of being confounded by non-validated and potentially problematic new assessment methods.

After reading the instructions and content article, the player arrives at the game's main interaction screen (see Figure 4.3). This screen is dominated by a blue workspace ("the sea"), with only the star-marked Causling Capital island present initially. The other islands start on a paginated "dock" near the top of the screen—the pagination effectively divides the islands into 6 groups, with the islands arranged alphabetically across them. Players drag the islands from the dock into the sea using the computer's mouse, positioning islands wherever desired. Islands can continue to be repositioned once in the sea, or even dragged back onto the dock. Once an island has been dropped into the sea, players can drag bridges out of the island's coast, stretching these bridges to reach other islands—the bridge is created once it is connected (and "snaps") to a second island. All bridges are initially marked with a "plus" arrow pointing in the direction the bridge was dragged, indicating an increasing causal relationship between the islands. Once a bridge has been created, the player can double-click on the arrow to change the bridge's causal direction (cycling through changes in both target direction and influence direction). Players can also right-click and remove a bridge that has been created. This simple set of interactions allows the player to easily construct a virtual causal map (in the form of bridges between islands) using the interactive system, in much the same way they may draw a causal map during a paper-based assessment. In game terms, players can choose islands for the Causlings to visit (i.e., to include in their causal map) and to build bridges between these islands (i.e., to specify directed causal relationships). The interactive system also displays a few just-in-time instructional tooltips about other interface elements (primarily based on the game framing, such as when the "Release the Causlings" button becomes enabled) in an attempt to help players learn how to play the game as they are playing.

Beyond enabling the core interaction of selecting and building bridges between islands to construct a virtual causal map, I implemented two different modes of gameplay (selected as an experimental condition, see Chapter 5) to explore the effects of different game dynamics on the assessment process. In the more basic version, which I will refer to as the "basic" mode, players build bridges and completely construct their causal map before clicking on a button labeled "Release the Causlings." Upon hitting this button, a number of Causlings (determined by the number of bridges connected to the first island) go racing out from the Causling Capital and attempt to cross all the bridges the players has built. If the bridge corresponds to a valid causal relationship, the Causlings successfully cross it and reach the next island. If the bridge does not correspond to a valid causal

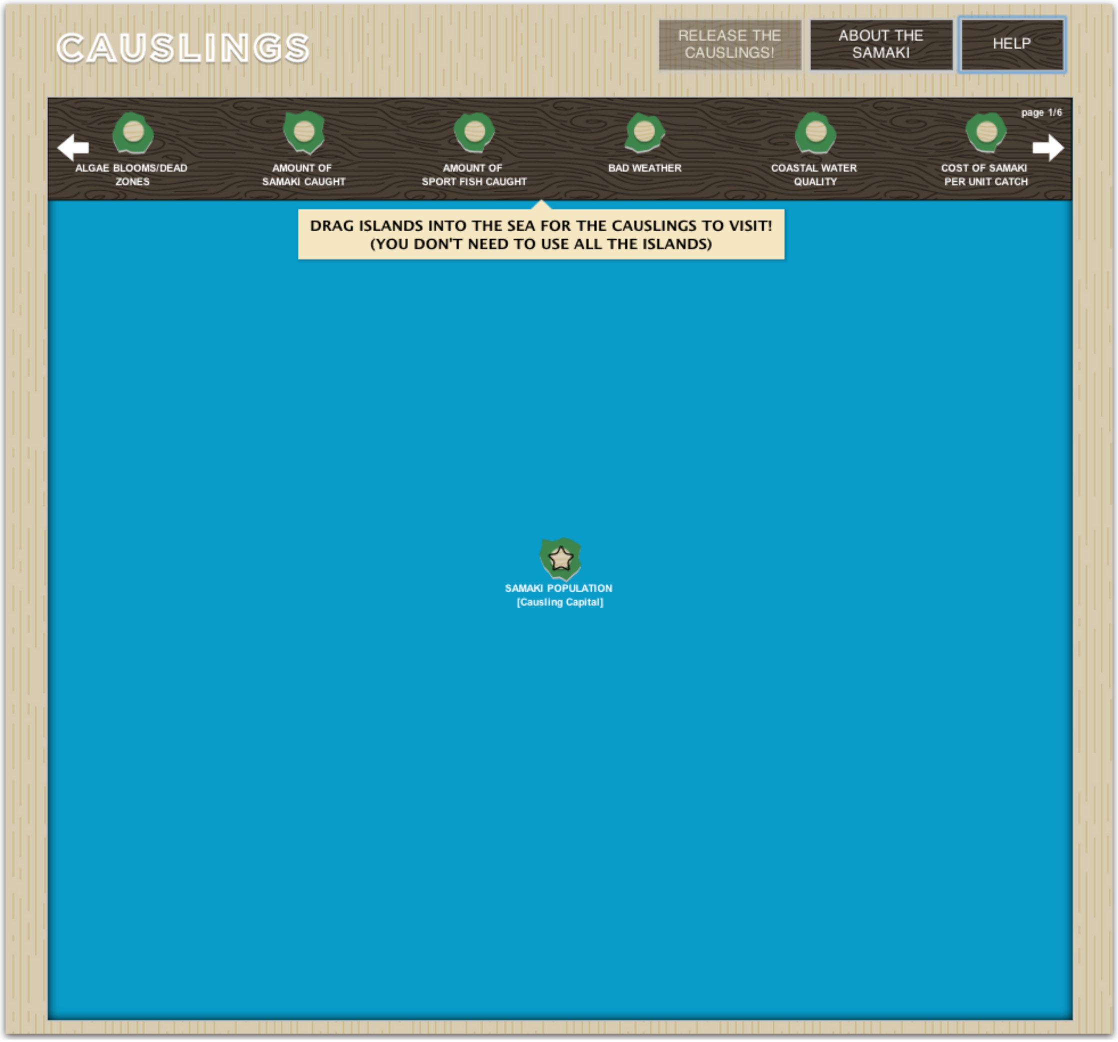


Figure 4.3: The *Causlings* game interface at the start of a new game. The help instruction tooltip disappears once the player begins playing.

relationship, the Causlings fall off the bridge and disappear into the sea. Once a single Causling has reached a new island, that island becomes settled with a small house appearing in the center, and more Causlings (again determined by the number of bridges connected to that island) begin to appear on that island and attempt to cross any connected bridges, including those previously traveled; the Causlings collectively perform a breadth-first traversal of the causal map in an attempt to visit all the connected islands. Once all the visited islands have produced new Causlings and all the Causlings have attempted to cross a bridge—either reaching their destination or falling into the sea—the game ends and a score is displayed. Thus in this basic mode, the player constructs a causal map using the interactive system (in almost the same manner as in Plate's assessment, though narratively framed as a game), and then immediately receives game-like feedback in the form of a short animation and a score revealing the accuracy of the constructed map. Note that while the Causlings are moving at the end of this game mode, the player is unable to alter the causal map.

The second game-play mode, referred to as the "timed" mode, attempts to make the system's construction causal maps even more game-like by introducing an additional game pattern—specifically the "Time Limit" pattern (Björk and Holopainen, 2005). In this mode, the game screen includes a timer counting down until the time that the Causlings begin to cross bridges and explore. Once the timer runs out, Causlings begin to appear at the Causling Capital, but one by one instead of all at once as in the basic mode. The Causlings also leave the island one by one, with a Causling periodically trying to cross a player-built bridge to a new island (see Figure 4.4). Once an island is reached and settled, more Causlings begin to appear on that island and begin to periodically leave in the same manner. The game is designed so that if there are too many Causlings on an island and not enough bridges, the Causlings will eventually move off the island and disappear into the sea, though because of timing issues (see below) this only occurs if a settled island has absolutely no connected bridges. Unlike in the basic mode, in the timed mode the player can continue moving island and building bridges while the Causlings are appearing and crossing bridges—the waiting time between when each Causling appears gives the player time to react and alter their map as desired, as well as to continue expanding the map.

In this way, the timed mode is designed as a kind of race, with players trying to construct a causal map before all the Causlings appear and try to cross the created bridges. Moreover, in the timed mode Causlings who begin to cross an incorrect bridge will "hesitate," changing color and pausing on that bridge before disappearing into the sea, thereby giving the player a chance to correct the bridge based on game feedback. The game ends once there are no more Causlings who wish to cross a bridge and explore; each island produces a number of Causlings based on the number

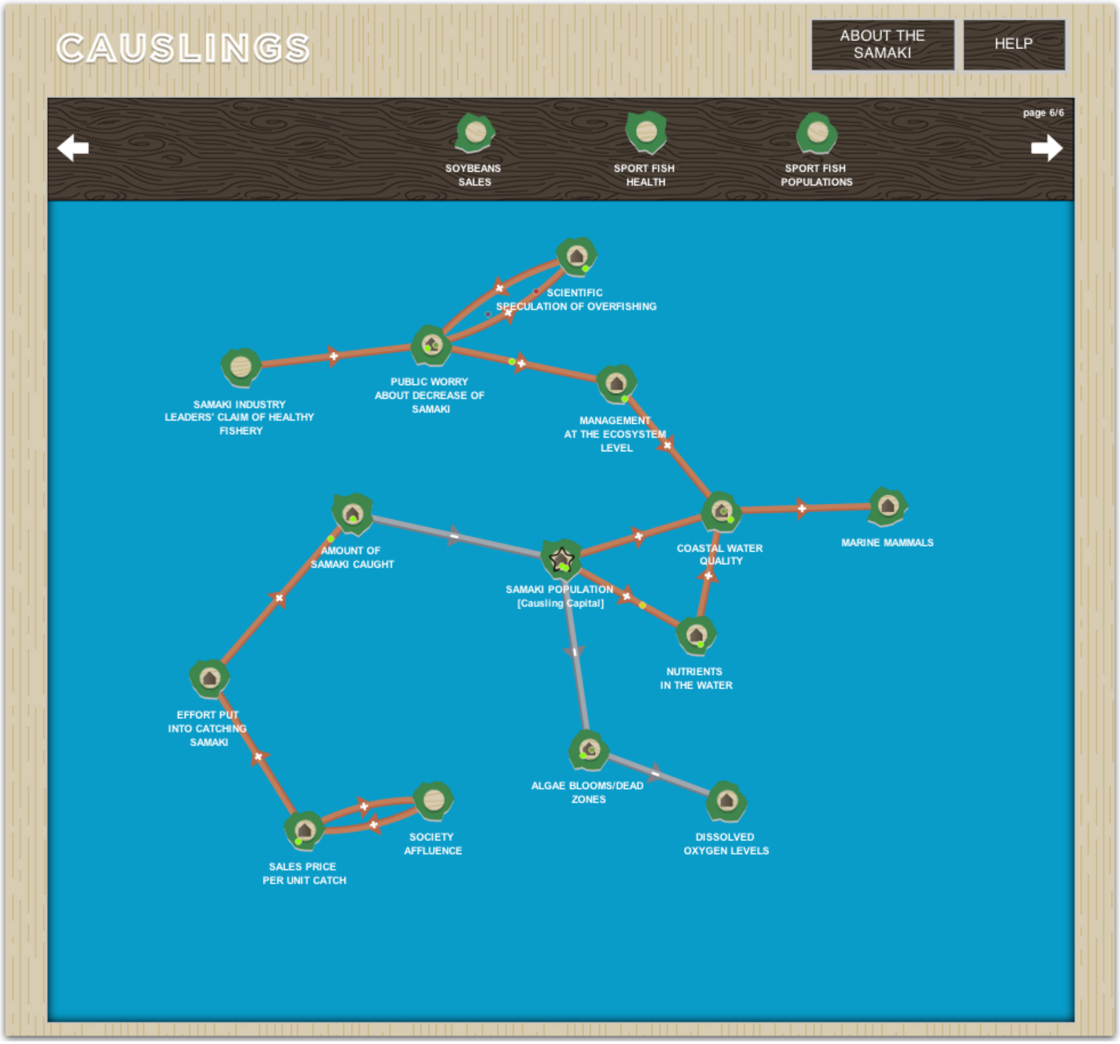


Figure 4.4: The *Causlings* game being played (by me) in timed mode. The bright green dots are Causlings waiting to move, the yellow dot is a hesitating Causling, and the red dots are the disappearing remains of fallen Causlings.

of connected bridges before it stops—though due to a bug in the game logic, an island that has stopped producing Causlings does not produce more if new bridges are connected to that island, causing games to end somewhat sooner than intended. Indeed, the timed mode required careful balancing of the timing (in how quickly Causlings are produced, how quickly they leave their starting island, how fast they move, etc.) in order to give players sufficient opportunity to construct their causal maps while maintaining an appropriate challenge to the building race. The timing likely had significant influence on player interactions and reactions to the timed mode—further design iterations may be necessary to properly balance the game's challenge (see Schell, 2008, for one discussion of game balance).

Overall, the timed game-play mode represents a model of interactive assessment in which players receives limited real-time feedback as to the accuracy of their maps—players can see Causlings fall off bridges, and then modify or remove those bridges as appropriate. With this variation I can explore the effects of the real-time assessment enabled by using an interactive computational system for measuring causality (e.g., on the process and results of causal map construction), as well as how an additional game pattern may effect the game-framing and players' engagement with the system.

In either mode, once the game is completed the player is given a "score" of how well they did in the game—in effect, the immediate results of the system's assessment of their causal understanding. This score includes a number of metrics: the player is told the number and percentage of islands visited (i.e., how many islands were validly connected to the Causling Capital), the number of percentage of Causlings settled, the mortality rate of the Causlings (the complement of the percentage settled), and a "final score" point value. This point value is calculated based on the final state of the map using the rubric developed by (Plate, 2006, see below), which considers the accuracy of each individual bridge at the end of the game. Thus while this score usually correlates with the other three (in that visiting more islands and settling more Causlings requires a greater number of accurate bridges), it is not a direct function of these measures. The game does not tell players the specifics of how this score is calculated in an attempt to maintain the game-like framing of the system, though the lack of score explanation deviates from the game design principle of offering feedback. Nevertheless, the existence of multiple metrics allows the player to in part choose their own method of measuring their success. After they receive this score, players have the option to play the game again using the same set of islands, starting over with only the Causling Capital island in the sea.

4.1.1 *Causlings* as Assessment

Causlings involves players in constructing a causal map, following previous work in assessing causal understanding (e.g., Plate, 2006, 2010; Shute et al., 2010). As this map is created virtually, it can immediately and automatically be scored, enabling players to receive a score (evaluation) at the end of each play attempt. The final score reported by the game provides an assessment of people's causal understanding of the complex system modeled by the game's content. This measurement is based on Plate's (2006) method for measuring the accuracy of a generated causal map by comparing a person's map to maps developed by experts in the domain[2]. In this method, the causal relationships defined by the subject—that is, the bridges built by the player—are compared piece-wise to those defined by experts, with each relationship given an accuracy value according to a rubric (see Table 4.1). This rubric gives higher values to causal relationships that are identified and agreed upon by multiple experts, with the assumption that agreement indicates a higher likelihood of truth or importance. If a causal relationship is not identified by any of the experts, it is given a negative value so that a purely random set of relationships has an overall slightly negative score. As Plate explains, a score from this system "can be interpreted as measure of the participant's success in identifying only those links that were also identified by the experts" (Plate, 2006)—players are not explicitly penalized for a lack of completeness (failing to identify a causal relationship identified by an expert), though less complete maps will be scored lower than more complete maps. However, in his studies Plate did not consider the influence direction in his analysis (though students did record influence direction on their maps), and it is unclear if the scoring balance of his rubric includes influence direction. Moreover, because of an implementation oversight on my part, incorrect bridges were scored at -0.5 points when they should have received -2 points (see below). Thus although the game's scoring is based on Plate's and uses a similar scale, scores from *Causlings* cannot be directly compared to score Plate's work.

[2] In fact, the prototype *Causlings* game uses the exact expert maps from Plate's work as a basis, so that this assessment matches prior research.

Table 4.1: Causality scoring rubric, originally developed by Plate. See below for details of "N^{th}-order" designation.

Number of experts identifying causal relationship	Points received
Four experts	8 points
Three experts	6 points
Two experts	4 points
One expert	1 points
Zero experts (1^{st}-order)	-0.5 points*
Zero experts (2^{nd}-order)	-2 points
Zero experts (3^{rd}-order)	-4 points
Zero experts (4^{th}-order)	-6 points
Zero experts (5^{th}-order)	- 8 points

* *score given to incorrect bridges in* Causlings.

Note that Plate's rubric also scores causal relationships that are only indirectly identified by experts—that is, causal relationships that represent a "zoomed-out" view of an expert's causal chain. These chains are identified by repeatedly multiplying the causal map's adjacency matrix by itself; thus causal chains with a single intermediary are found in the "2^{nd}-order matrix" (the adjacency matrix squared), causal chains with two intermediaries are found in the 3^{rd}-order matrix (the adjacency matrix cubed), and so forth. Plate's rubric includes up to 5^{th}-order matrices, thereby including a safety margin beyond the four-step chain that was "the longest chain that [he] could identify that could reasonably be contracted into one step" (Plate, 2006). However, in developing the *Causlings* game I found anecdotally that validating anything beyond a 2^{nd}-order matrix makes the game seem too easy. With higher order matrices, randomly placed bridges are likely to be at least partially accurate, so that while the overall final score may be equivalent (being counter-weighted by larger penalties for incorrect edges), the player was less likely to see the feedback of Causlings falling off an incorrect bridge[3]. Yet only scoring based on 1^{st}-order matrices (matrices representing direct

[3]I did consider using a probabilistic function to determine if a Causling successfully crossed a bridge, so that higher scored bridges (such as those that were agrees upon by more experts, or even those that were more direct) would allow more Causlings to cross. However, I felt that the apparent randomness of whether a Causling survived a crossing appeared unfair. Because success in the game draws heavily from a player's potentially flawed mental model of the complex system—indeed, the game is designed to reify and assess this model—it is important to be as clear as possible about how the game reacts to that model so as not to seem arbitrary and thus not meaningful (see Salen and Zimmerman, 2003).

connections identified by experts) seemed to miss some valid connections identified by other experts (researchers in the field of environmental science) who play-tested the game; thus *Causlings* uses 2^{nd}-order matrices to measure the player's bridges' accuracy. In addition, the final scores reported to the players, while calculated using the above rubric, are multiplied by a factor of 10 in order to better match the type of score expected in a video game (e.g., getting a score of 130 points is more genre-appropriate than getting a score of 13 points). However, for clarity and ease of comparison with other research, all score measurements reported in this dissertation follow the base rubric values.

As discussed in Chapter 2, there are potential flaws in using a comparison with experts as a basis for assessing the accuracy of a causal map (as well as arguments against cognitive mapping as an assessment of understanding in the first place). For example, the expert maps are themselves constructed through a cognitive mapping process, and thus reflect the assumptions, beliefs, and biases of their creators—there may be only weak associations between the expert map and an external reality. Though this limitation is alleviated somewhat by the use of multiple experts, collective misunderstanding of a complex system is still possible. Nevertheless, expert maps provide a method for concretely measuring the accuracy of causal maps that can be effectively integrated into an interactive computer system such as *Causlings*. Furthermore, as the expert comparison method has been documented in prior research, adopting this method allows me to more concretely discuss the effects of this computer- and game-based assessments in comparison to other work (e.g., previous paper-based assessments performed by Plate.) Thus despite its limitations, comparison to experts currently provides the most feasible method for measuring the accuracy of causal maps in order to assess causal understanding through an interactive computer system.

So in sum, playing the *Causlings* game involves following the same general procedure as Plate's CMAST methodology for assessing causal understanding. Like in the structured form of 3CM (Kearney and Kaplan, 1997) used by Plate, players choose which causal factors to include in their causal map by dragging the islands representing those factors off of the dock and into the sea.

Players then identify causal relationships between the chosen factors by building bridges between the islands—in much the same manner as a person might draw a line between causal factors in a paper-based assessment. In this way, interaction with the system is simply a computerized version of Plate's assessment with an added game narrative. The timed gameplay mode does deviates somewhat from this established technique for assessing causal understanding in order to make the cognitive mapping process more engaging, but it is the nature of this deviation that is under investigation in this dissertation. Thus *Causlings* is an interactive computer system and game that seeks to assess causal understanding using the cognitive causal mapping technique established by prior research. This computational assessment system is able to provide a faster measurement by immediately calculating the accuracy of a causal map, and has the ability to gather further data on the *process* by which players construct their causal maps.

4.1.2 Collecting Map Construction Data

The *Causlings* interactive system prototype is able to collect data about the process through which subjects construct causal maps. These data are primarily related to the spatial organization of the causal maps (e.g., what islands were positioned where[4]), and the sequence of map construction (e.g., what bridges were created when). Specifically, the system records "game moves"—significant actions taken on the part of the player to influence the game system (see e.g., Salen and Zimmerman, 2003; Lindley, 2005)—with the intention of considering the construction sequence and being able to reproduce maps created by subjects during user studies. Thus the system records events:

1. When a new game is started by a particular player (with each player given a unique random id number upon first accessing the *Causlings* systems). Note that each player may start multiple new games if subsequent plays are attempted.

[4]This is in contrast to the graph-based view of a causal map, as mathematical graphs do not have a particular planar layout.

2. When an island (factor) is moved on the screen—either from the dock or to a different spot in the sea. The system records the new (x,y) coordinates of the island relative to the application screen.
3. When a new bridge (relationship) is successfully built between two islands. The system records the target direction (which bridge was the source and which the target); the influence direction is "increasing" by default.
4. When the causal direction of a bridge is changed. This can be either the target direction or the influence direction, as both are toggled through the same interaction.
5. When a bridge is deleted by the player.
6. When an available modal dialog displaying the samaki article was opened and closed.
7. When the Causlings begin to explore (in timed mode, this event occurs during bridge building).
8. When the Causlings finish exploring (e.g., at the end of the game). This event also records the final score and statistics of the game as presented to the player.

Note that each of these events is recorded with a timestamp with millisecond precision, as well as the logging id of the player. The data for all game attempts was written to file on the system's game server for future analysis.

Overall, these data allow me to reconstruct the overall process by which players create causal maps using the system. However, to avoid significantly increasing processing overhead or overcomplicating the recorded data, there are a number of potentially valid measures that are not recorded by the prototype interactive system. For example, the system fails to record when players scrolled through the the "dock" of islands, which could indicate how the player interacts with the pre-supplied list of factors (e.g., considering them in order, searching for a particular factor, etc). Moreover, the system does not record individual mouse movements, which could be used to identify factors or relationships that the player explicitly choose *not* to include—the player might hover the mouse over an island while deciding whether to select it, before moving on. Similarly, the system only records the successful creation of new bridges, omitting events where the player starts to create a bridge from an island, but then perhaps changes his or her mind and so does not complete the bridge. In these ways, the *Causlings* prototype interactive system is able to record explicit map construction events, but has little quantitative data on choices or objects that players rejected in the process of reifying a causal map.

4.2 Implementation Details

As with the design of any interactive system game (or indeed any game or artifact), developing *Causlings* required countless design decisions, many of which may have been made unconsciously or based on personal taste. In this section I describe the history of some of the more salient and relevant decisions, discussing the motivation behind certain design choices—particularly ones that involved changes from established methods of assessing causal understanding. In this way, I will validate my design decisions and provide grounding for the design of future related systems.

4.2.1 Architecture

Causlings is a web-based game written almost entirely in JavaScript, making use of a number of 3rd-party libraries and plugins (such as the popular jQuery library and its extensions)[5]. In particular, the graphical interaction is implemented through the Raphaël.js[6] library for drawing vector graphics—this library allows me to use the Scalable Vector Graphics (SVG) specification for rendering both simple shapes and complex paths (such as the islands' coasts), as well as providing DOM-level access to the drawn shapes so that they can be manipulated through common JavaScript methods like those used by jQuery. Within the JavaScript implementation, islands and player-

[5]The source code for the *Causlings* game is part of the Causality Project, and can be viewed at `http://github.com/highbulp/wc_master`. The bulk of the game code can be found in the `public/javascripts/game.js` file.
[6]Available at `http://raphaeljs.com/`.

created bridges are stored in associative arrays of JavaScript objects, with keyed references to an associative array of Raphaël.js Element objects that represent what is actually drawn on the screen. In this way, the graphical representation is separated from but accessible to the underlying data structures, following a model-view-controller pattern. The JavaScript application is served from a Ruby on Rails backend, though the game interaction is processed purely on the client-side—the Rails framework is used only for handling server requests for stand-alone pages such as the introductory narrative, and for logging game interaction events that are recorded by the client.

The *Causlings* game uses this particular implementation because of the game's organic growth as an extension of the Causality Project (Tomlinson and Black, 2011a,b). The Causality Project uses a Ruby on Rails implementation to store an online database of "issues" (causal factors) and "relationships" (causal relationships, particularly with specified causal direction), and then to allow a user to browse chains of these relationships. In developing a graphical visualization of these causal relationships and the complex web they form (i.e., the causal map), the Causality team and I choose to use the Raphaël.js library for its ease of use; it allowed for rapid implementation and integration with the existing Rails system, as well as enabling features such as animations for user interactions. The *Causlings* game is based on this initial visualization codebase, extending it to support a wider variety of user interactions such as building and modifying bridges. Note that while the game was rapidly developed and can be run on a wide variety of platforms (including some mobile devices), differences in browser implementations of the JavaScript engine can have deleterious effects on the prototype's robustness and performance across platforms.

4.2.2 Prototype System Content

Mirroring Plate's (2006) prior research, the causal factors that form the islands' controlling concepts are drawn from the complex system surrounding the overfishing of a fictional species of fish known as the samaki (an alias for the menhaden species). According to the article presented to players, the samaki plays an important role in the ocean ecosystem by controlling algae growth through feeding and acting as a food source for other species of sport fish. However, evidence suggests that the samaki may be in danger of being overfished due to increase demand by human industries—yet this evidence itself may be influenced by other external factors (e.g., drops in the price of samaki may have led to decreased fishery profits, rather than declining species populations). The causal factor "Samaki Population" acts as the Causling Capital (where the Causlings begin their explorations), as this factor was the most highly connected in the expert maps and offers a clear central concept around which the controversy can be framed. Other causal factors include "Algae Blooms/Dead Zones," "Cost of samaki per unit catch", and "Public worry about decrease of samaki"—in total there are 36 causal factors and subsequent islands.

The use of this complex system is borrowed directly from Plate's work, and thus shares some of his assumptions and design motivations: for example, samaki is used as an alias in order to "ensure that prior knowledge about the specific issue would not be a factor" (Plate, 2006). Indeed all the causal factors used in the game are borrowed from Plate's work, and are based specifically on a short article written by Plate about the controversy surrounding the samaki (see Plate, 2006, for details). Using this article as the only source of prior knowledge makes playing the game in some respects a measurement of reading comprehension rather than an assessment of existing causal knowledge. This limitation was appropriate for Plate's research, as he used the cognitive mapping process to assess systems thinking skills—how linear are the maps users make when given a set of content, and how is this influenced by systems thinking training? On the other hand, my goal with the *Causlings* game is to assess how accurately players understand a causal system. Such an assessment would theoretically be applied to externally gathered knowledge to answer questions such as "how well does the subject understand the causality of climate change" (and indeed, such an application is important future work, as described in Chapter 9). Nevertheless, using this previously developed article as the sole basis of players' causal knowledge allows me to normalize the evaluation of this prototype game on a particular set of knowledge, enabling better comparisons of the *Causlings* prototype system with prior, non-computer-based assessment techniques.

4.2.3 Game Framing and Narrative

In designing the game framing for *Causlings*, I drew significant inspiration from the classic computer game *Lemmings*. In *Lemmings*, the player guides a population of oblivious virtual creatures past a number of lethal obstacles (long falls, fiery pits, etc.) by commanding individual creatures to perform actions such as climb a wall, dig through the earth, build a short staircase, or simply "block" to hold back the ever-increasing crowd. Winning the game primarily involves solving the puzzle of determining what sequence of commands to give the creatures to let them forge a path past the obstacles, while also requiring the player to deal with the careful timing of when to give the Lemmings commands as they march to their otherwise inevitable doom. *Lemmings* has been a highly popular game since its release (versions are ported to new gaming systems to this day), and is one of my personal favorite video games. Thus in presenting *Causlings* as a game I seek to harness some of the enjoyable and addictive qualities of this classic game.

Indeed, *Lemmings'* narrative goal of "find a path on which these creatures can safely move" seemed a natural fit to my initial considerations of making a game about constructing causal paths or chains. Early project designs involved having the player define relationships between causal factors to try and form a causal chain between two indirectly related factors (e.g., the player would need to build a path of accurate causal relationships from "take the bus" to "climate change"). In a way, such a design would resemble a causality-based version of *Wiki Golf*, a game in which the player tries to navigate between two random Wikipedia pages by following as few links as possible. While this kind of path-finding game could have enabled exploratory learning (e.g., players could try and discover unexpected causal relationships and indirect effects), I choose to focus on the assessment of causal understanding as a more significant contribution than simply supporting the discovery of causal chains—particularly as focusing on such linear causal chains can actually contribute to a *lack* of causal understanding and systems thinking (see e.g., Grotzer and Perkins, 2000) by instead emphasizing reductionist and linear thinking.

Nevertheless, the exploration metaphor at the root of a path-finding game seemed appropriate for supporting the construction of causal maps, following previous cognitive mapping techniques for assessment. By shifting the "travel the path" framing to a "grow the territory" framing, I am able to support non-linear map building through the exploration metaphor—rather than having the player find a specific path to a particular destination, the player constructs multiple branching paths to reach a multitude of locations. Refining this narrative, I developed the metaphor of treating causal factors as islands and causal relationships as bridges linking the islands, making a game about exploring islands by building bridges. I considered alternative exploration metaphors, such as digging tunnels between caves, but the island metaphor[7] easily supported narratives of failing to cross an inaccurate bridge, as well as having thematic resonance with the overfishing controversy that provides the content for the prototype game. Thus the game's narrative works to encourage the player to construct as large and as complete causal maps as possible, with the goal of visiting every island.

The *Causlings* interactive system deviates somewhat from the CMAST framework by allowing players to select causal factors to include throughout the process of building the map, rather than ahead of time (following Plate, players do select factors from a pre-determined list). Furthermore, players are able to continuously reposition the islands included in their map, making their own groupings and layouts to help them make sense of the model they are illustrating (within the limitations of the computer's screen real estate—the game's interface can get crowded once there are a lot of islands and bridges in the sea, and providing sufficient space for map construction was a key concern, particularly for smaller displays). Although I considered automating dynamic island layout, such as through force-directed layout algorithms (e.g., Kobourov, 2007) or even giving islands assigned starting positions in the sea, allowing players to spatially organize their own layouts

[7]Serendipitously, the island metaphor also provides a nice homage to the EcoRaft project (Tomlinson et al., 2009), one of the first research projects I worked on in graduate school.

can be suggestive of how they think about or construct these causal maps, as well as matching to previous subject-driven cognitive mapping techniques. Thus the interaction and gameplay of the *Causlings* system gives the player control over the construction of their causal map, though the framing narrative has potentially significant influence on the cognitive causal mapping.

4.2.4 Appearance

In designing the look and feel of *Causlings*, I (along with assistant undergraduate designers) aimed for an aesthetic that had an overall "crafty" but stylized feeling (see Figure 4.3 and Figure 4.4). The islands were based on scraps of colored construction paper, with circular wooden chips at the center. The dock, borders, and buttons are also intended to be reminiscent of wooden boards—yet only reminiscent, as the grain-like pattern in the wood is stylized rather than realistic. The bridges are simple lines (a previous textured overlay made them look like wooden sticks, but had to be removed to improve the interaction frame-rate). Similarly, the Causlings themselves are simple colored circles—though colored in a way to look like painted wooden chips—to enable faster rendering. This craft-based design aesthetic matches the environmental themes of the game, with the stylization following from the project's scope as a research prototype. However, it is possible that the lack of photo-realistc or 3D graphics may have influenced player's reactions to the system, making it seem less "game-like" (through lack of polish) and more of a testing system. Such is an accurate assessment of the project's status, but has potential implications for how the system was received and interacted with.

As a tool for visualizing causal relationships, *Causlings* uses the form of a concept map, rather than alternative structures described in Chapter 2, such as Ishikawa diagrams or Growing Polygons. The game uses colors and iconography to indicate the influence direction of constructed bridges—increasing bridges are orange and marked with a "plus" sign, while decreasing bridges are blue-grey and marked with a "minus" sign. Arrows in the center of the bridges indicate target direction. These colors were chosen to match the aesthetic, while also lacking strong connotations in the domain of sustainability. Size is not used as a visualization technique, as the game does not consider causal strength and there is a lack of screen real estate to support larger visual elements. Similarly, position is left up to the player to determine, rather than explicitly conveying an aspect of a causal relationship. (It is worth noting that in the timed mode, the manner in which the interaction is implemented causes short bridges between closer islands to require slightly less time for Causlings to cross, but this difference is slight enough that it is unlikely to have a noticeable effect on the game or players' interaction with the system).

Thus, the visual appearance of the *Causlings* prototype follows from the CMAST framework on which the game is based, with aesthetic choices made to support automating the process of cognitive causal mapping.

4.3 Conclusion

In this chapter I have described *Causlings*, a prototype interactive system and game for assessing causal understanding. This system enables users to reify through a digital system causal maps that can be used in assessing their understanding of an underlying causal system, following the CMAST causal mapping technique of eliciting user models of the causality of complex systems. The computerized system is able to record effectively the process by which a player constructs a causal map, allowing me to consider the sequence by which subjects construct causal maps and how that sequence might inform an assessment of causal understanding. The system also adds a game framing to the CMAST technique, introducing an exploration narrative to causal mapping: players define causal relationships between causal factors as bridges for the Causlings to cross and explore, making the bridges as accurate as possible so the Causlings do not fall off. Gamifying this assessment technique has the potential to increase user engagement, thereby possibly encouraging a greater effort in participation and enabling the measurement of collective understanding that can better support the development of new environmentally sustainable policies and actions. Yet strategic play introduced by such gamification could influence the construction process; this dissertation research explores the nature of that influence.

In the next chapter, I introduce an evaluative user study designed to test the *Causlings* system's ability to explore the process by which players perform cognitive causal mapping, as well as to test the effectiveness of the game framing in developing this assessment tool.

Chapter 5. Evaluating *Causlings*

This chapter describes the experimental design of a user study performed to evaluate and test the developed prototype of the *Causlings* system. Participants in the study played the *Causlings* game, constructing causal maps that I then analyzed for insight into players' systems thinking. Players also provided feedback about the experiences, which I use to illustrate some effects of the system's game framing on assessment.

This evaluation seeks to provided grounded evidence in support of answering this dissertation's overarching research questions presented in Chapter 1, namely:

1. What can interactive systems reveal about the structure of causal maps and the process by which people construct these maps as reifications of their understanding of complex causal systems?
2. What characteristics of games as a form of interactive systems support or impede assessment, particularly of causal understanding?

I address the first question by exploring the causal maps that players produce through playing *Causlings*. While I consider the form of the final, completed map in a manner similar to prior research involving cognitive mapping (i.e., in terms of visual layout, graph connectivity, etc), my primarily focus is on the sequence of player actions taken to construct those maps (i.e., what factors were included and what bridges were built, and in what order). I analyze the maps constructed in this user study for patterns in spatial layout, map creation sequence, and for other revealing anomalies. This evaluation thus acts as an initial exploratory, quantitative study to consider the process by which people construct causal maps, as revealed by an interactive computer system.

This evaluation addresses the second question by considering the effects of the prototype's game-based design on its effectiveness as simultaneously an assessment and a game. This study evaluates the level of engagement (one of the main benefits of games for education) fostered by the system—that is, how successfully this computerized assessment functions as a game while continuing to assess causal understanding. At the same time, I compare the system's assessment of player's causal understanding across the basic and timed game modes as an exploration vector into how utilized game patterns may affect the system's capabilities as an assessment tool. With this evidence as a starting point, I seek to probe the suitability of games such as *Causlings* for supporting stand-alone assessments.

To summarize, this study and its analysis is developed around answering the following evaluative questions, which are described in more detail in the discussion of my analysis methodology below:

1. How does the *Causlings* interactive system function as an assessment tool?
2. What forms do the causal maps constructed by players take?
3. What patterns can be found in players' sequence of map construction?
4. What characteristics make *Causlings* effective or ineffective as an engaging game?

In answering the first question, I consider how *Causlings* functions as an assessment, considering whether particular design components (e.g., game patterns) influenced the assessment goal that is at the core of the interactive system. With the second question, I consider the results of the cognitive mapping performed in this system along the lines considered by prior research (e.g., Eden et al., 1992; Plate, 2006) by looking at the structure of completed maps, though with a particular emphasis on spatial layout. The third question offers an initial foray into the underexplored topic of the process people use to create cognitive causal maps, looking at the sequence of map construction. Finally, the fourth question investigates the efficacy of some of the benefits of gamifying cognitive mapping suggested by previous literature—specifically the level of engagement afforded by gamified systems—as well as considering other impacts of framing the interactive system as a game.

The rest of this chapter describes in detail the evaluative user study. I begin by describing the participants recruited for this study, and then explain the research activities that they performed and what data was collected from these activities. I end with an explanation of the analysis methodology of the collected data, which will frame the presentation of the study results in the next chapter.

5.1 Study Participants

In order to evaluate the *Causlings* system with as large and diverse a population as possible, I recruited study participants from three different pools of subjects: students from a local high school, undergraduate students at the University of California, Irvine, and anonymous participants recruited through the Amazon Mechanical Turk (MTurk) crowdsourcing marketplace. These different populations allowed me to mirror the populations in studies by Plate: both our studies include undergraduate participants at large research universities, as well as younger students (though my study involves high school students as opposed to Plate's middle school students). Furthermore, recruiting participants from MTurk allowed me to evaluate how players from settings outside of academia interacted with the system, gaining a better view of how the cognitive causal mapping of interest to this dissertation might occur "in the wild."

No matter the recruitment method, each participant was asked to fill out a short demographic survey before beginning the user study. This survey asked participants to report their age, gender, and to answer short questions to measure their familiarity with computers and video games, as well as their attitude towards the environment and sustainable issues.

Computer Usage Participants were asked: *"Are you good at using computers?"*, with response options of "Yes", "I'm okay I guess", "No", or "I never use them" (valued 4 to 1 respectively on Table 5.1). This question aimed to account for how comfortable subjects would be interacting with the computer system.

Gaming Frequency Participants were asked: *"Do you play computer games or video games?"* with response options of "Yes, I play them all the time", "Yes, I play them sometimes", "I rarely play them", or "No, I never play them" (valued 4 to 1 respectively on Table 5.1). This question sought to measure familiarity with video game patterns, as well as indicating potential attitude towards games (i.e., people who play video games frequently may be more amenable to the forms of engagement offered by games).

Attitude toward Environment Participants were asked: *"How important are the environment and environmental issues to you (e.g., pollution, climate change)?"* and responded on a five-point Likert scale ranging from "Very important" to "Very unimportant" (valued 5 to 1 respectively on Table 5.1). This question was intended to account for how likely subjects would be engaged in the environmental subject content of the *Causlings* game.

The results of this survey are detailed in Table 5.1; in addition, each of these subject populations is described in more detail in the following subsections.

Table 5.1: Summary of user study participants.

	High School students	**University students**	**Mechanical Turkers**	**All Subjects**
Number of Subjects	26	34	29	89
Average Age	17.1 years (s.d. 0.80)	21.1 years (s.d. 3.04)	29.1 years (s.d. 8.15)	22.6 years (s.d. 6.95)
Gender	19 male 7 female	13 male 21 female	11 male 18 female	43 male 46 female
Computer Usage	3.7 of 4 (s.d. 0.49)	3.3 of 4 (s.d. 0.46)	3.9 of 4 (s.d. 0.35)	3.6 of 4 (s.d. 0.50)
Gaming Frequency	3.2 of 4 (s.d. 0.86)	2.5 of 4 (s.d. 1.05)	3.2 of 4 (s.d. 0.64)	2.9 of 4 (s.d. 0.91)
Attitude toward Environment	4.2 of 5 (s.d. 0.59)	4.0 of 5 (s.d. 0.76)	4.0 of 5 (s.d. 0.87)	4.1 of 5 (s.d. 0.75)

Note that the use of different subject pools means that each group had very slight variations in study procedures (e.g., depending on whether the user study occurred in-person or online). These variations are detailed in the next section along with the study procedures. To reduce these variations, I structured participant recruiting to have as much in-person interaction and observation of subjects as possible (as with the high school and university students), so that I could better control events and handle any problems that arose. However, I forwent this requirement with the MTurk group in order to gather as large a subject pool as possible—particularly as recruitment across groups occurred simultaneously when the number of participants was in question. In the end, this variety of subjects offers a broader view of how players may interact with the *Causlings* game system.

Because of the relatively small number of participants from each subject pool, much of the time I do not differentiate between subject pools in my analysis of the study results—the subject's recruitment pool is not always a significant variable in my analysis. While there may be notable differences between subjects based on how they were recruited, the subject pools are too small to determine significant differences dependent on pool (not to mention that not all data was gathered for each subject pool, as described below). My research questions do not focus on how different groups interact with *Causlings*, but instead consider how a range of subjects interact with the game system. The exploratory nature of this study (as well as time and scope limitations) leads me to consider the results as a whole, rather than dividing the results based on their respective recruitment source.

5.1.1 High School Students

The first group of study participants was recruited from a suburban high school in Southern California. This high school has a distinguished academic program, as well as an overall diverse student body. Subjects were juniors and seniors in a four-year honors IT elective program that focuses on teaching topics such as introductory programming, everyday computer use, and computer game development—indeed, many of the participants had experience creating their own digital games. Overall, subjects from this group were expected to be highly familiar with computer systems and games (and so likely to have well-developed systems thinking skills), while also representing diverse backgrounds.

In total, 26 students from this subject pool participated in this user study. These participants were 16 to 18 years old (average 17.1), with 7 of 26 female students. As expected, 2/3 of the students reported being skilled at using computers, with the other 1/3 reporting being "okay", and on average participants reported playing video games "sometimes" (5 of 26 reported playing games "rarely"). These participants were familiar with computers and video games, and as such likely amenable to the design and use of a game framing in *Causlings*. Furthermore, participants responded to the question "How important are the environment and environmental issues to you (e.g., pollution, climate change)?" with an average of 4.2 on a Likert scale (with 5 being "very important" and 1 being "very unimportant"). This response suggests that the group would likely be engaged with the environmental content—that is, the issue of samaki overfishing—that forms the causal system that is mapped through the game prototype. So overall, these students could be expected to be well disposed towards the interactive game system.

Through arrangement with the instructor of the honors IT program, the user study was performed in a classroom visit during the hour normally devoted to the elective. Students under 18 had previously received permission from their parents to participate in the research, and all students had been given the samaki article the night before to read ahead of time (so that as much time as possible could be spent playing *Causlings* during the visit). The evaluation was conducted within a computer lab with which the participants were highly familiar; students choose seats from about 6 rows of computer workstations facing forward. The computers in the lab appeared to be relatively older machines running Windows 7, though one student used his own laptop to participate. As this study took place during the week following a string of standardized tests (both state-based testing and Advanced Placement tests), the students seemed accustomed to alterations in routine and had a slightly relaxed, informal attitude about participation. I quickly introduced the user study (informing students that they would be playing a game and filling out a survey about their experiences), and then gave the students the URL for the system and let them explore and play on their own. This introduce was as brief as possible (as well as reiterated through the IRB-required study sheet) so as not to bias any results.

The students then accessed the *Causlings* game on the lab computers and began the study procedures, while I watch for any problems or difficulties. Because 26 students were participating all at once, I was unable to take highly specific notes about their interaction, and as such I rely primarily on the logged data to track interactions rather than my own observations (indeed, the quantitative logged data is of much higher validity than qualitative data in general; see below). During the study I mainly focused on making sure everyone was able to play the game during this first public deployment. And in fact, the version of the Firefox browser[1] initially used by the students ended up not being able to support the JavaScript-based animation of the Causlings—when the system tried to release the Causlings, either because the timer expired in timed mode or the player hit the "Release" button, nothing happened. This error was discovered about 20 minutes into the hour. Luckily, we were able to have students update the browser on their individual machines and restart the game with enough time left to complete the procedure. Nevertheless, this error makes the logged data somewhat messy; yet while the interaction is like less representative of how a full game session would proceed, the process by which students construct their causal maps remains valid with or without animated Causlings. In effect, all students played the "basic" mode, without any designated stopping point, before eventually playing a full game. In addition, because of another system error influenced by this required restart, only 7 of 26 participants completed the closing evaluation survey (described below), and these surveys could not be reliably linked to game-play logs. Because of these initial errors, there is a lack of summative data from this sub-

[1]The student participants were asked to play *Causlings* using Firefox as a browser, as the version of Internet Explorer available on the lab machines had been tested as unable to handle some of the JavaScript libraries needed to render the game's graphics. *Causlings* had been tested against an old version of Firefox (v3.0.4)—but apparently not old enough.

ject group, though there is a sufficient quantity of data on gameplay process (i.e., bridge-building) to support analysis of the sequence and results of the computer-based causal mapping that is of primary interest to this dissertation.

5.1.2 University Students

I recruited the second group of study participants from the University of California, Irvine (UCI): a large research university in Southern California. Participants were primarily undergraduates, recruited through the Human Subjects Lab (HSL)—a program run through the school of Social Sciences that gives undergraduate students the opportunity to participate in ongoing research as subjects of minimal-risk research experiments. Students are offered extra credit in select courses for participating in these experiments (at a standard rate per time spent), and sign up to participate through an online portal—students self-select for the experiments in which they wish to partake. In total, 34 students were recruited through this system. These students averaged 21.1 years of age, 2/3 of them were female, they were familiar with using computers (2/3 reported being "okay"), were much *less* likely to play games than the other groups (more than half played "rarely" or "never"), but on average still viewed environmental issues as "somewhat important". I did not inquire as to major or field of study, as such information was not also relevant for participants from the other subject pools.

These students performed the study procedures individually in an academic office setting on the university campus. I gave each participant a brief introduction similar to the one I gave to the high school students, and then they played the *Causlings* game on a desktop computer running Mac OS X with a large, high-resolution screen; the size of this screen resolved problems concerning screen real-estate, but the increased pixels-per-inch did add some difficulty in selecting particular portions of each island to either drag or build bridges. The game was played on the latest version of the Google Chrome browser, on which it had been primarily developed and tested (anecdotally, Chrome's JavaScript engine maintained the highest level of responsiveness while playing *Causlings*). While participants played the game, I observed from over their shoulders, recording general notes about their play behaviors to supplement analysis from gathered game logs.

5.1.3 Mechanical Turkers

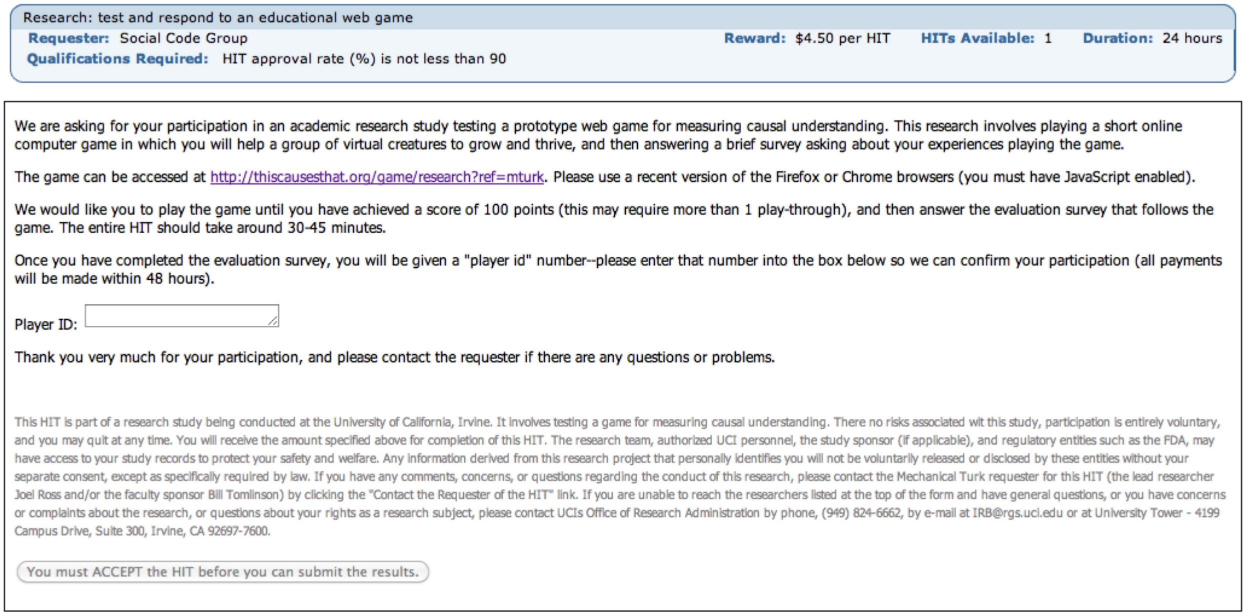
The final group of study participants was recruited through the MTurk online crowdsourcing system. MTurk allows users to distribute cognitive work to a large number of laborers, with the work broken down into simple, one-time tasks that workers are paid to complete. These tasks are most commonly human computation problems: tasks that are difficult for computers and yet simple for humans (e.g., image labeling). Through MTurk, requesters create Human Intelligence Tasks, or HITs, specifying the job and the amount paid for its completion—usually ranging from as little as $0.01 for a quick task up to $10 dollars for more involved jobs, such transcribing audio clips. Workers, referred to as "Turkers," who log into the MTurk website are able to choose which tasks they perform (after previewing the HIT), forming a micro-task marketplace (Kittur et al., 2008). MTurk was launched in 2005, and as of 2010 Amazon reported that the system had more than 400,000 registered workers, with about 50,000 to 100,000 HITs to work on at any given time.

Turkers remain almost entirely anonymous to requesters, who are shown only a randomly generated ID number to represent individual workers. However, previous research (e.g., Ipeirotis, 2010; Ross et al., 2010b) has explored the general demographics of Turkers, as self-reported through surveys posted as HITs on the system. This research found that while around half of Turkers are from the U.S., 35% are from India. 2/3 of U.S. Turkers are female, while 2/3 of Indian Turkers are male; Turkers tend to be younger (25 to 35 years old), well educated (most have some college education or a Bachelor's degree), and of mixed socio-economic status as determined by reported income (with the income of Indian Turkers being substantially less than that of U.S. Turkers). The 30 respondents who participated in this user study reported ages from 19 to 49 year old, with an average of 29, and to be a group skilled at using computers (an average of 3.9 out of 4, as expected from those engaged in clickwork) who played video games sometimes (only 3 reported playing video games rarely) and who on average saw environmental issues as "somewhat important" (though 3 considered environmental issues "somewhat unimportant"). As I used the same data gathering instruments as the other two subject groups (who are known to be in the U.S.), I did not survey Turker participants about their nationality; that said, the written responses suggest that at least one Turker participant was not a native English speaker.

Turkers participated in this user study remotely, locating and accepting a HIT posted on MTurk (shown in Figure 5.1). Participants were asked to play the *Causlings* game until they achieved a score of 100 points (or a "10" on the rubric described in Chapter 4), which in testing was commonly achieved on either the first or second play-through. This requirement ensured that participants put effort into playing the game, rather than simply creating random bridges; such assurances are necessary because of how the extrinsic payment offered through MTurk can crowd-out intrinsic motivation, as well as to protect against fraudulent workers (see e.g., Kittur et al., 2008; Silberman et al., 2010; Downs et al., 2010, for further discussion of fraud on MTurk)[2] After completing the closing survey, Turkers were provided a unique number to enter back into the MTurk system, allowing me to link anonymous accounts with their survey and log data, and thereby compensate each worker. Turkers were paid $4.50 as compensation for their participation—as the study took around 30 to 45 minutes, this allowed for a fair pay rate of $6-9/hour.

[2] An aside: much of Amazon's description and writing about MTurk works to frame the system as a black-boxed cloud service—requesters submit their hard computational problems to the service, and they come back as solved within a few days. This framing "mechanomorphizes" (Caporael, 1986) the human workers into simple components in a larger mechanical system; posting a HIT to MTurk is making a "remote person call" (Bederson and Quinn, 2011). Indeed, a significant portion of my own MTurk research (e.g., Ross et al., 2010b; Silberman et al., 2010) aimed to dispel this potentially dehumanizing framing.

In the University of California, using MTurk for research followed this mechanomorphic view: using funds to pay research subjects was equivalent to using funds to pay for a cloud service, such as remote storage or computer processing (the computers in this case just happened to be people). Yet in the process of recruiting subjects for this user study, I discovered that the university accounting offices had begun restricting funding used to pay Turkers due to various liability and hiring issues; the UC system was starting to consider Turkers as freelance workers rather than as machines. In this way, the arguments others and I made in past research almost thwarted my recruitment of subjects for this current research!



Research: test and respond to an educational web game

Requester: Social Code Group **Reward:** $4.50 per HIT **HITs Available:** 1 **Duration:** 24 hours

Qualifications Required: HIT approval rate (%) is not less than 90

We are asking for your participation in an academic research study testing a prototype web game for measuring causal understanding. This research involves playing a short online computer game in which you will help a group of virtual creatures to grow and thrive, and then answering a brief survey asking about your experiences playing the game.

The game can be accessed at http://thiscausesthat.org/game/research?ref=mturk. Please use a recent version of the Firefox or Chrome browsers (you must have JavaScript enabled).

We would like you to play the game until you have achieved a score of 100 points (this may require more than 1 play-through), and then answer the evaluation survey that follows the game. The entire HIT should take around 30-45 minutes.

Once you have completed the evaluation survey, you will be given a "player id" number--please enter that number into the box below so we can confirm your participation (all payments will be made within 48 hours).

Player ID:

Thank you very much for your participation, and please contact the requester if there are any questions or problems.

This HIT is part of a research study being conducted at the University of California, Irvine. It involves testing a game for measuring causal understanding. There no risks associated wit this study, participation is entirely voluntary, and you may quit at any time. You will receive the amount specified above for completion of this HIT. The research team, authorized UCI personnel, the study sponsor (if applicable), and regulatory entities such as the FDA, may have access to your study records to protect your safety and welfare. Any information derived from this research project that personally identifies you will not be voluntarily released or disclosed by these entities without your separate consent, except as specifically required by law. If you have any comments, concerns, or questions regarding the conduct of this research, please contact the Mechanical Turk requester for this HIT (the lead researcher Joel Ross and/or the faculty sponsor Bill Tomlinson) by clicking the "Contact the Requester of the HIT" link. If you are unable to reach the researchers listed at the top of the form and have general questions, or you have concerns or complaints about the research, or questions about your rights as a research subject, please contact UCIs Office of Research Administration by phone, (949) 824-6662, by e-mail at IRB@rgs.uci.edu or at University Tower - 4199 Campus Drive, Suite 300, Irvine, CA 92697-7600.

You must ACCEPT the HIT before you can submit the results.

Figure 5.1: The MTurk HIT, as seen by Turker participants.

5.2 Study Procedures

As alluded to above, the overall structure of the user study to explore the process of causal mapping as revealed by the *Causlings* prototype system involved participants playing the game (often multiple times), and then responding to their play experiences through a closing electronic survey and, where applicable, in an interview with me. Data from this user study was collected in the form of electronic logs of game interactions (see Section 4.1.2), observations I documented as the facilitator, audio-recorded interviews, and the responses to the closing survey. Taken together, these data allow me to analyze the process by which players construct causal maps, as well as how players interacted with and reacted to the game system.

5.2.1 Playing *Causlings*

After agreeing to participate in the user study (per IRB requirements) and filling out the demographic survey, players were presented with the online *Causlings* game system as described in

Chapter 4. Players read through the introduction and instructions, as well as the article describing the samaki. This reading generally took around 5 to 10 minutes—a number of subjects ended up skimming the samaki article in order to get to the game, making their exposure likely equivalent to that in Plate's studies even though they were able to review the article as needed. Upon reaching the main interaction screen, players would begin moving islands and building bridges, proceeding until the game ended (either because they released the Causlings in the basic mode, or because time ran out in the timed mode). When the game ended, players were given their score and the options either play again or to take the closing survey.

Participants were asked to play multiple games, either playing until the (about) half-hour of time allotted for the study ran out (in the in-person studies, as with the high school and university students) or until they had achieved a sufficiently high score (in the online studies performed by Turkers). As players sometimes needed more than one attempt to understand fully how the game was played, this repetition helped assure that the causal maps players constructed were as accurate as possible, as well as enabling me to collect a greater quantity of observations. A single run through the game generally took between 5 minutes (in timed mode) and 15 minutes (in basic mode). Differences between intermediate and final play-throughs are discussed in Chapter 6, Section 6.1.

Gameplay mode—either basic or timed—was assigned randomly to each participant upon first accessing the interactive system. This random assignment supported a between-subjects experimental design, allowing me to compare the two game modes (e.g., in terms of reported engagement)[3]. Note that the method of random assignments means that I did not force a balanced distribution (each participant had a 50% chance of playing either mode), so there may be an unequal number of players in each condition. Once a game mode was assigned, a participant played that same game mode on repeated play-throughs (within the same gaming session), so as to avoid confusion and to support the experimental design. However, early user studies—specifically, the ones involving

[3]During the development of *Causlings*, playtesters found the timed mode to be subjectively more difficult, and so suggested offering it as an "advanced mode" option. Such a choice could be implemented for a future deployment, but was not appropriate for this user study experiment.

high school students and early studies with university students—encountered a programming error in which subsequent plays either from refreshing the page or hitting the "play again" button would switch the game into basic mode, if it were not already. As such, I do not include the results of these studies in my analysis comparing assessments in the two game-play modes (see below), though I do include them in my consideration of the game as a whole, particularly when using the results of these play-throughs in analyzing players' construction of causal maps. Similarly, other unexpected JavasScript errors occasionally but rarely forced participants to restart the game; these abandoned games are not considered in my analysis—only fully completed games (i.e., games where a final score was electronically logged) are included.

5.2.2 In-Person Observations and Interviews

During the user studies performed in-person, I manually recorded observations of the players and their interactions with the *Causlings* game. For the high school student user study, this observation was highly casual—I was mostly watching for program bugs or other problems using the system during this initial test by 26 simultaneous participants. As such, I only recorded general observations about the group's mood and reaction towards the game, as well as getting an anecdotal sense of what kinds of maps would be constructed by first-time users as opposed to my increasingly experienced play-testers. In the one-on-one user studies with university students, however, I took much more careful notes about each individual's specific interactions with the game. In particular, I watched for general patterns in island positioning and bridge building (though I relied on the electronic game logs to record the individual, specific moves)—for example, I noted whether players dragged multiple islands into the sea before connecting them with bridges, as well as how those islands seemed to be organized. I also strove to record the players' verbal asides and reactions (comments either to me or to themselves, gestures, and smiles or laughter). With these notes, I aimed to begin forming a broad understanding of some of the patterns of cognitive causal mapping that could guide my quantitative analysis of the game logs, as well as to collect reactions to supplement responses to the closing survey. In this way, this observational data is used to guide and supplement the more carefully collected electronic data—the qualitative data is illustrative, but not conclusive.

The in-person user studies also concluded with a short (usually 5 to 10 minute), semi-structured interview. With the high school students, this took the form of a group interview focused more on a critique of *Causlings* as a game, as the students themselves were studying game design. Due to time constraints and the group interview format, I was unable to collect more detailed qualitative responses regarding how the players constructed causal maps. With the university students, I asked questions mirroring those in the closing survey (see below), as well as about particular interactions I had observed—for example, if the player built their map in a certain pattern, reviewed the samaki article frequently, or made an interesting comment. Thus I asked questions such as:

> *"Tell me about playing this game."*
>
> *"How did you choose which islands to include in your map? How did you choose which bridges to build?"*
>
> *"Was there any reasoning in how you organized islands on the screen?"*
>
> *"What did you change about how you played that your score improved (or worsened) in later games?"*
>
> *"What did you mean by a particular comment?"*
>
> *"How did you decide when you were done building bridges and to release the Causlings?"* [in basic mode]

All interviews (for both high school and university students) were audio recorded for review and analysis.

Note that there is a high probability that the responses to these interviews (and potentially the closing surveys, described below, with in-person user studies) were affected by the social dynamics introduced by the structure of the interview. Subjects may have been likely to assume that I as the person running the experiment created the system they were interacting with[4] and evaluating,

[4]Indeed, a number of subjects asked me—usually at the close of the interview—if I built the game.

which may have influenced their reactions to the system. Players may have been more or less careful in their map construction than if they were not being observed (though subjects were not aware that game moves were recorded), and were likely less willing to offer critiques of the system's framing as a game. For this reason, the collected qualitative data resulting from surveys and interviews is not the main source of my analysis, but rather is used to help illustrate and inform a quantitative analysis based on the electronically recorded system interactions.

5.2.3 Closing Survey

After playing the *Causlings* game the requested number of times, players took a short closing survey to evaluate the system's framing as a game. The survey began with a number of open-ended questions asking the players to reflect on their experience playing the game:

1. Please describe your strategy for playing the game. How did you choose what bridges (causal connections) to include or exclude from your map? How well did your strategy work?
2. Overall, what did you enjoy most about this game? Is there anything you would change about this game to make it more interesting?
3. Overall, do you think you learned anything by playing this game? If so, what? What could be changed about this game to help you understand these causal connections better?
4. After playing this game, do you feel that samaki are in danger of being overexploited? Please explain your answer. If you do not feel informed enough to answer, what else would you want to know about this issue in order to form your own opinion about what should be done with the samaki?

Question 1 aimed to solicit player's own views of the process they used to construct their causal maps, framed within the context of the game. Moreover, by framing the process in game terms as a strategy being used, with this question I hoped to interrogate how the act of "playing a game" may have influenced the cognitive mapping process. Question 2 gathered subjective opinions about how well *Causlings* functioned as a game, focusing on specific elements or patterns that made the game more or less enjoyable. Question 3 asked for a subjective self-assessment of whether any learning occurred through playing *Causlings*. While the primary purpose of this system is to assess causal understanding (over enabling further learning about causal systems), with this question I aim to discover potential ways in which future versions of the game may also support learning and systems thinking. Question 4 was borrowed directly from Plate's (2006) study procedures, and is asked as a kind of summative assessment to validate the player's consideration of the samaki's ecosystem. I use the responses to this question to inform my analysis of how well players "understood" the game, and as a point of comparison to the constructed causal maps. In addition to this open-ended question, players were directly asked: *"How well do you feel you understand the causal relationships that are part of the game?"* (with possible responses "Very well," "Pretty well," "Somewhat," and "Not at all")—this question was also intended to act as a kind of "sanity check" if players' self-assessment of their causal understanding matches with the results from the game's use as an assessment tool.

The player was then asked a series of 5-point Likert-scale questions intended to measure their level of engagement in playing the *Causlings* game (see Table 6.13). In line with previous research (e.g., Sweetser and Wyeth, 2005; Brockmyer et al., 2009; Fu et al., 2009), I consider the level of engagement with a game primarily in terms of the game's ability to support a state of flow (Csikszentmihalyi, 1991; see Chapter 3 for further discussion). The questions themselves were drawn from the EGameFlow questionnaire (Fu et al., 2009) for measuring engagement in educational games, which itself is adapted from the GameFlow model (Sweetser and Wyeth, 2005). EGameFlow is a validated model for assessing flow (which most of the game design scholars in whose work I ground my own research—e.g., Salen and Zimmerman, 2003; Juul, 2005; Schell, 2008—accept as a measurement of engagement), and has been applied to the kinds of educational games that are related to *Causlings*, though my research focuses on assessment. Thus this questionnaire is an appropriate tool for measuring engagement, one of the main potential benefits of gamifying an assessment. In order to keep the length of the survey as short as possible and thus to get the most considered responses (as well as keep the time requirements of the user study reasonable), I selected representative questions from the EGameFlow questionnaire to use in this study. These questions were selected based on my own understanding of the GameFlow model, as well as questions that had analogs in the Game Engagement Questionnaire (Brockmyer et al., 2009), another common game evaluation measure. The player responses to these questions (though they may suffer from some biases) forms the basis for my analysis of Causlings' success as a game—that is, whether and in what form does it offer an engaging experience while also assessing causal understanding.

5.3 Analysis Methodology

As described above, I collected a wide variety of data through the study procedures that form this evaluative user study, such as system interaction logs, survey responses, and interview content. I use these data in a mixed-methods approach towards answering the research questions guiding this dissertation (though again, the qualitative data is primarily used to illustrate the quantitative). In this section, I detail the process by which I use these data to explore the causal mapping process and evaluate the *Causlings* game. This section is organized around the four evaluative questions introduced at the beginning of the chapter, with the questions divided into subsections. In each subsection I discuss which data are applicable to the particular question, and what tests and analyses are utilized to answer that question. The results of the user study presented in Chapter 6 will also follow this organizational structure.

5.3.1 How does the *Causlings* interactive system function as an assessment tool?

The first evaluative question considers how the *Causlings* game prototype functions as an assessment of causal understanding. While I consider the accuracy of the constructed causal maps, my analysis is not a robust evaluation of the accuracy of the game's assessment. Creating accurate educational assessments is a difficult, unsolved problem (Bloom et al., 1956; Gipps, 1994; Shulman, 1986). Designing an assessment requires some existing observation or measurement of a student's skill or ability that can be used as a basis for measurement (e.g., is the math problem solved correctly, or does the essay make a sufficient argument?). Assessment design is further complicated when the item being assessed is internal to the subject, such as cognitive processes like "understanding." Designing a novel, valid, and reliable assessment for determining a person's internal understanding of causal relationships is outside the scope of this dissertation; thus I rely on the existing assessment procedure of causal mapping (specifically in the CMAST form used by Plate) to act as a kind of "ground truth," despite the limitations of this technique discussed in Chapter 2. Yet because these scores are ungrounded, my analysis focuses not only on how high the scores are, but also the map components that led to those scores: e.g., which constructed bridges were correct or incorrect.

Furthermore, in order to begin considering how the interactive system (and its game framing in particular) may influence the assessment process, I compare the assessment results of both game modes: basic and timed. While interaction with the basic mode represents the "default" use of the system, the timed mode introduces an additional game pattern while also suggesting how the desirable game quality of regular feedback may influence the assessment process. In short, I consider whether the game modes produce similar measures of causal understanding. I compare the assessments performed in the two game modes through the primary measurement of causal understanding: the final game score based on the rubric developed by Plate for measuring the accuracy of a causal map in comparison to experts (described in Chapter 4). As most players played multiple games, I consider the score of the final game they played, as well as for the highest scoring game and the average game score across all play-throughs. I compare these measurements through a comparison of means (via Student's t-test) of the basic and timed game modes. I hypothesize that the timed mode and basic mode will produce equivalent measures of causal understanding, suggesting that the additions of the timed mode do not significantly influence the reliability of the interactive system's assessment. Furthermore, I use responses to the survey question *"How well do you feel you understand the causal relationships that are part of the game?"* to evaluate if the game results match the player's own self-assessment, checking for correlation between the response to this question and the final scores of players from all three subject pools whose final evaluations were received. Again, this measurement is compared between the basic and timed game modes in order to evaluate any potential differences introduced by the change in game patterns.

Additionally, I analyze the completed causal maps looking for commonly created bridges that are either correct or incorrect according to the expert models. This analysis can help to identify potential patterns of failure that may be influenced by the interactive system (e.g., were certain types of relationships harder to consider when timed?), as well as common points of confusion in the experiment itself (e.g., what parts of the article were systematically unclear?). Indeed, a subjective consideration of how "easy" or obvious the mistakes were, such as in comparison to Plate's students, can also inform the potential validity of this tool for assessment—did one mode of game-play cause more or different simple mistakes than the other? Overall, I quantitatively use the range and components of game scores to examine the effects of the interactive system (and its variations) on assessment, in comparison to Plate's original results, in order to measure the system's utility.

In this section of the analysis I also briefly examine the related question of whether the *Causlings* game fortuitously supported learning about causal relationships. I primarily analyze player's self-reported learning through responses to the closing survey's open-ended question using a qualitative process and presenting representative examples. While this analysis is in no way intended to offer robust proof of learning, it should indicate potential learning benefits that may be harnessed in future designs of similar interactive systems were the goal of creating a stand-alone assessment dismissed.

To reiterate, this component of my analysis focuses on comparing the game modes to each other in order to examine any effects that particular design choices (such as game patterns) have on assessment, rather than robustly validating the *Causlings* game as an assessment technique. I rely on the prior research upon which this game is based to form a valid assessment method, with my own research exploring variations on that procedure.

5.3.2 What forms do the causal maps constructed by players take?

The second evaluative question begins my examination of the forms of causal maps players construct during *Causlings*, with a focus on the form of the finished map at the end of a game (consideration of the steps taken to construct this final map are addressed with the third evaluation question in the next subsection). This question asks: "what kinds of maps do players make using this interactive system?" Considering the structure and content of these maps can help both reveal how the *Causlings* system functions as an assessment—does the game encourage or support creating particular types of causal maps—as well as further exploration of what the structure of a reified causal map reveals about causal understanding. In addition, this form of analysis is comparable to how causal maps are studied in prior research (e.g., Eden et al., 1992; Markóczy and Goldberg, 1995; Plate, 2006), helping to contextualize the maps created with this game in terms of previous work on cognitive causal mapping. Note that throughout this analysis I compare the maps constructed during the basic and timed game modes, as well as to the results reported by Plate when appropriate.

I start by analyzing the causal factors and relationships that players chose to include in their final causal map—that is, what islands they used and what bridges they built. Following Plate's example, I look at usage counts to identify which islands and relationships are commonly used (incorrect or not). This analysis is supplemented by players' own subjective views of what islands they used and why, as solicited through interviews and the open-ended questions in the closing survey. Furthermore, because this dissertation engages with the problem of visually representing causality, I also examine the visual structure that players give to the causal maps they construct. I study the visual organization and spatial layout of constructed maps, including a number of examples of maps that were created by players playing *Causlings*. Such consideration is supported by the computational basis of the *Causlings* system. Again, this analysis of the visual structure and organization is supplemented by observational data and player responses to interviews and surveys, allowing me to explore both intentional and unintentional patterns of map organization.

In this section, I also analyze the maps from a graph theory perspective. Causal maps can be trivially converted into adjacency matrices: the entry in the i^{th} row and j^{th} column represents the direction of the causal bridges (if any) from island i to island j. Such matrices allow me to consider a number of simple graph theory metrics to support understanding the structure of these maps. For example, I look at: the number of nodes (islands) and edges (bridges) included; the ratio between these two, also known as the link density; the degree of each node—that is, the number of edges pointing into the node (heads) and the number pointing out (tails), which can indicate which nodes are more "central" to the graph; and the lengths of the longest chains and sequences to measure how spread out the graph is. The last two items in particular can let me measure the overall "shape". According to Eden et al.:

> "In principle when a shape is flat (lots of short paths between tails and heads) it may indicate little depth of thinking but contrarily it also suggests consideration of a high range of choice and alternative views. A thin tall shape (a small number of long paths between tails and heads) may indicate detailed argument without a consideration for alternative definitions of the situation" (Eden et al., 1992).

Thus these measurements can help to explore the graph's status as a representation of causal understanding, and can be contrasted with the final scores given by game's rubric. Finally, I also measure the graph's using Plate's Web-like Causality Index (WCI, see Plate, 2006)—as my research significantly builds off of his, I make sure to address his primary item of interest (the non-linearity of causal maps). However, the game framing of *Causlings* instructs players to "visit as many islands as possible" leaving from a single starting island, implying a need for branching rather than linear maps (e.g., in order to get from 1 place to many places, you'll need to branch); as such I expect maps generated by this game to be more web-like than those reported by Plate.

Overall, looking at the final structure of the maps in terms of their visual and graph-based organization (an analysis enabled by the presented interactive computational system) can help inform an assessment of causal understanding, as well as possibly point to some of the effects of the system's game framing.

5.3.3 What patterns can be found in players' sequence of map construction?

With the third evaluative question, I begin to consider the *process* by which people construct causal maps, looking beyond a consideration of only the final produced map. I am interested in the sequence of steps people take in order to graphically represent the causal understanding that is in their heads, as well as how the construction process might influence the maps that are generated. With this dissertation, I perform an initial exploration of the sequence and process of cognitive causal mapping, looking for patterns that can begin to shed light on how people construct causal maps and on how that process is related to their understanding of the causal relationships that make up complex systems. As the process of building causal maps is currently understudied in previous literature, my initial consideration involves a wide exploration of a large set of qualitative data (interaction logs, observations, interviews, etc.) that can begin to identify interesting aspects for further study. Because this question is somewhat broadly framed, I direct my exploratory analysis of the gathered data with a number of sub-questions and hypotheses concerning the process of constructing causal maps (though they may raise further questions along the way). To aid in organization, each of these questions is described separately below.

5.3.3.1 Do players select islands or create bridges in a particular order?

The first question considers the ordering of the sequence of map elements included (i.e., game moves a player takes) in the construction of a causal map. Are certain bridges created earlier than others, or are there other patterns to the order that bridges are built (e.g., bridge Y is usually built shortly after bridge X)? This ordering may be indicative of how players understand causal relationships and the causal mapping process. For example: players may build bridges earlier because they are the most "obvious" or are otherwise central to how the player thinks through the causal relationships in the presented system. Moreover, such patterns may coincide with players' causal understanding as measured by the final game score—players who understand the causal system better may be more likely to build bridges in a particular order.

I explore this question primarily by analyzing collected game log data. I parse out the sequence of islands used and bridges built, calculating which islands and bridges are most likely to be used first. This quantitative data analysis is supplemented by observational data (i.e., patterns I noticed in watching players play the game), as well as any interview and survey responses that might speak to this ordering. In a way, this analysis extends the consideration of islands and edges included described in Section 5.3.3 and originally performed by Plate; I now also account for the ordering of island and bridge inclusion.

5.3.3.2 Do players construct causal maps hierarchically?

In exploring players' sequences of bridge building, I focus on the patterns organizing the construction of causal maps. In particular, I hypothesize that the game's design choice of having Causlings begin at a single island may cause players to construct causal maps in a hierarchical fashion—players start with connections made to the Causling Capital, and then branch out from there to include more indirect relationships. In this way, players construct maps using a kind of breadth-first traversal, building the map out to have further and further reach. Alternatively, players may use a more depth-first traversal and build bridges in sequences of independent chains starting at the Causling Capital—following each chain as far as they can before beginning another. I hypothesize that one or both of these patterns may be common because of how people might systematically try and solve the "puzzle" of building an accurate causal map. Moreover, the prominence of linear thinking (see Section 2.2.2) suggests that it is likely that people may construct causal maps by making linear chains representing the various linear causal stories they can tell, and then connecting those chains.

I look for these hierarchical building patterns in analyzing the collected interaction log data. I calculate the frequency in which players either build out from the center (breadth-first hierarchies) or construct long individual chains (depth-first hierarchies) in order to explore how common these patterns may be, as well as what relationship they have to causal understanding.

5.3.3.3 Do players have conscious plans or strategies for constructing causal maps?

Last, I interrogate to what extent *Causlings* players are self-aware of their own process of constructing causal maps. Do they pre-plan a particular organization for their maps, or follow a conscious heuristic or algorithm for building? Moreover, as *Causlings* gamifies the cognitive mapping process, how do players thinking about map building as a series of game moves—do they form any short-term tactics or long-term strategies to "win" the game by constructing causal maps in a particular game? While the simple design of *Causlings* as a game does not necessarily require complex play strategies to win, I am interested in whether players use any such strategies. Indeed, I hypothesize that different game-play modes may encourage players to focus on using particular tactics to try to win the game, thereby altering the process by which they construct causal maps (in terms of sequence of steps taken, as identified above).

As this question considers player's self-awareness, I test my hypothesis primarily through players' qualitative responses to direct survey and interview questions. The first open-ended question in the closing survey speaks directly to this topic, and I repeated this question during interviews (with a focus on the particular game moves and map building I observed). I discuss and give examples of the results of these probes. Nevertheless, due to the potential for response bias, these examples are intended to be illustrative and evocative, rather than demonstrative.

5.3.4 What characteristics make *Causlings* effective or ineffective as an engaging game?

The fourth and final evaluative question considers the effectiveness of the *Causlings* prototype as an engaging and enjoyable game. As discussed in Chapter 3, gamification has the potential to increase engagement with systems for learning and assessment—a highly engaging tool for measuring causal understanding could support more accurate assessments by encouraging subjects to try harder, or could even support large-scale analysis of the public's (mis-) comprehension of the causal relationships in topics surrounding sustainability. In order to determine what aspects of the game can support assessment in this manner, I explore what components influence the interactive systems's level of engagement and players' overall reactions to the game's design. In short, I look at whether the game supports participation in assessment, and what aspects of the game influence this support.

Engagement with the game system is measured primarily through the EGameFlow-based Likert-scale questions asked at the end of the closing survey. Player responses are aggregated and analyzed in order to consider what aspects of flow may or may not be encouraged by this game prototype. In addition to an overall measure, I compare the player evaluations between the basic mode and the timed mode of the game. In this way, I can explore the specific effects of this implementation of a Time Limit game pattern on supporting engagement in a cognitive mapping game. I highlight this quantitative measurement of engagement with qualitative data gathered from the open-ended closing survey questions, interview responses, and my own observations. I identify particular moments that are either highly engaging or off-putting to the player, and use these events to point to characteristics of the game that may support or hinder assessment. For example, I discuss events that provoke smiles or laughter from the player, as well as events or design decisions that lead to confusion. Overall, with this evaluative question I explore how players interact with the game prototype, and how engaged they are in using this assessment tool.

5.4 Conclusion

In this chapter I have described the experimental design of a user study to evaluate and test the *Causlings* prototype interactive system and game. Through this procedure, I am able to study the forms of causal maps that are constructed through playing, the sequence in which those causal maps are constructed, and how players engage with the system and its game framing. While this study does not aim to validate the use of causal maps as an assessment technique beyond its support in the literature (e.g., Plate, 2010), it allows me to begin a novel exploration of the process and sequence of causal mapping as revealed by this computational system by looking at users' step-by-step construction of these causal maps. Furthermore, with this study I explore how how particular design components–such as the use of game dynamics—may influence the causal maps that are created, and thus influence the assessment of causal understanding. In these ways, the study procedures described here are appropriate for answering this dissertation's research questions and supporting this research's contributions to knowledge.

In the next chapter, I present the results of this user study and the analysis thereof, organized by the evaluative questions described above.

Chapter 6. Results and Analysis

This chapter presents and analyzes the results of the user study evaluating and testing the *Causlings* system prototype described in the previous chapter.

As described in Chapter 4 and Chapter 5, subjects in the user study used the interactive system in one of two variations of *Causlings* (either the basic mode or the timed mode), assigned randomly. The breakdown of the random distribution is detailed in Table 6.1. Players in all conditions played multiple games; my analysis includes all completed games (games which were given a final score at the end), or a total of 333 games across both conditions and all subject pools. Note that as explained in Chapter 5, subjects from the pool of high school student encountered a programming error, and thus did not play a consistent game mode across these multiple games (these games were also not correctly linked to the 7 closing evaluations collected, so data from the evaluations is considered as being from a "mixed mode" condition). This bug was also not resolved for the first university student subject, who played a single game in timed mode and further games in basic mode; this student is counted as part of the basic mode condition (the mode for the majority of her games).

Although programmatically random, the distribution ended up favoring the basic mode of interaction in terms of players (with 25% more participants) and number of games (with 20% more plays). Nevertheless, these populations are both sufficiently large to be compared in my analysis. The differences in the average and median number of games played across subject pools and game modes are most likely artifacts of the experimental procedures. I had asked Turkers to play until they

Table 6.1: Distribution of game modes among user study subjects.

	High School students	University students	Mechanical Turkers	All Subjects
TIMED MODE				
Num. players	*n/a*	19	16	35
Num. games	31	94	56	182
Median games per player	*n/a*	4	3.5	3
BASIC MODE				
Num. players	*n/a*	15	13	28
Num. games	39	62	51	151
Median games per player	*n/a*	4	2	3
BOTH MODES				
Num. players	26	34	29	89
Num. games	70	156	107	333
Median games per player	*n/a*	4	2	3

reached a particular score, while students in the high school and university student groups played for a set amount of time—thus potentially playing more games than it took them to reach the same score. Similarly, the timed mode enforced a particular length of game, while in the basic mode players could choose how long each game lasted (by choosing when to release the Causlings). Thus within the amount of time given for the user study, subjects in the basic mode condition were able to and did play more games on average.

I must note that, based on observations and comments, some (6-8) of the subjects in this user study seemed to misunderstand the objective of the *Causlings* game—likely because of the broad scope and complexity of the game's instructions and presentation (see below). For example, a few players did not understand the abstraction of the island's controlling concepts, and so tried to lead the Causlings to islands that would provide some benefit to the virtual creatures. As one player explained: "I tried to first make sure to include a healthy lifestyle to grow the causlings with the help of the human population then strategize by making a system where I could reproduce and sell the fish in a consistent way" (P9, University student). Because of the large amount of introductory reading, other players mixed up the Causlings and the samaki, and so played as if the samaki were the creatures moving from island to island (and as such would want to go to islands such as "Nutrients in the water" directly from the Causling Capital). A number of players also were confused about the concept of causal direction—a phenomenon I discuss in more detail in Section 6.2. Despite these apparent misunderstandings, I include these subjects in my analysis. Because of the assessment's framing as a game, players still attempted to understand and beat the game, even if they were not aware of the underlying logic behind their moves. Such unguided interaction is likely in the deployment of any interactive system or game, and thus can continue to inform the effectiveness of using such an interactive system for assessment of potential wide populations.

Following the structure of Chapter 5, the results of this user study are organized in terms of four central evaluative questions that consider respectively: (1) how the interactive system functions as an assessment tool; (2) the forms of causal maps constructed by *Causlings* players; (3) the sequence by which players construct these causal maps; and (4) the characteristics that make *Causlings* effective or ineffective as an engaging game. Each subsection of the chapter addresses a particular question—within these subsections, I present aggregated data from the user study relevant to the topic, and then analyze how those data support an answer to the section's guiding question. This approach will help to organize my analysis of the large quantity of data collected in this user study to best support this dissertation's research questions and contributions.

6.1 How does the *Causlings* interactive system function as an assessment tool?

I begin considering the how the interaction system functions as an assessment tool by looking at the scores achieved in the game across different game modes, detailed in Table 6.2. Theoretically, by the time they played their final game, players should have developed the best understanding of both how to play the game and of the causal system being mapped and thus would achieve the highest score of all their games. However, almost 25% of players got a lower score on their final game than they did on at least one previous play-through. For some of these players, their overall game scores were regularly low, suggesting that they did not fully understand the system (and thus the last game was simply another trial that may have gone more poorly than previous attempts). Other players, particularly from the University student pool, may have continue to experiment with the system after achieving their highest score, and simply run out of time allotted to the user study before they could construct another high-scoring map. Indeed, almost all Turker players stopped when they achieved their high score—a score sufficient to complete the HIT. However, the highest score achieved may not be representative of the players' overall causal understanding, as some players may have earned a high score by accident (as suggested by a low average score or final score). As such, Table 6.2 includes the average scores in each of these three categories. In addition, Figure 6.1 shows the trend in changes in game scores over multiple games for both game modes (measured as a function of the number of games played). This figure demonstrates that the majority of players improved their score over the course of playing *Causlings*—the slope of the linear regression for each player's games was 7.4 points (s.d. 12.9, median 3.6), demonstrating that on average players improved their score by around 7 points each game (though with a very high amount of variance). Thus players generally—but not always—improved their score over repeated plays, either from better understanding the interactive systems's representation of causality, altering their maps based on feedback from previous runs, or even potentially improving their understanding of the causal system.

Across all three categories, there was no significant difference between the scores achieved by players of the basic mode and players of the timed mode. This finding suggests that adding the Timed game pattern to *Causlings* does not significantly affect the system's ability to assess causal understanding. Although not significantly different, on average players in the timed mode did achieve higher High Scores and Last Scores (though very slightly lower average scores) than players in the basic mode. This difference is potentially due to the feedback provided in the timed mode—

Table 6.2: Differences in game scores across game modes and subject groups. "Average Score" is the average score achieve by the player across the multiple games played, "High Score" is the result of the highest scoring game, and "Last Score" is the score of the final game played by the subject. There were no significant differences between game modes.

	Average Score	Average High Score	Average Last Score	Avg. Report Understand	Corr. Avg. Score w/ Rep. Undst.
BASIC MODE					
High school	19.68	*n/a*	*n/a*	*n/a*	*n/a*
University	9.63	21.79	15.53	2.26	0.098
Turkers	12.21	17.75	14.75	2.81	0.201
Combined	10.81	19.94	17.29	2.51	0.166
TIMED MODE					
High school	23.54	*n/a*	*n/a*	*n/a*	*n/a*
University	9.00	21.87	14.8	2.13	0.194
Turkers	11.93	22.31	22.08	3.00	0.378 *($p < .05$)*
Combined	10.36	22.07	21.3	2.54	0.194
BOTH MODES					
High school	21.83	*n/a*	*n/a*	*n/a*	*n/a*
University	9.35	21.82	15.21	2.21	0.141
Turkers	12.09	19.79	18.03	2.90	0.219
Combined	10.61	20.89	19.31	2.52	0.184
p-*values of differences between modes*					
High school	0.18	*n/a*	*n/a*	*n/a*	
University	0.44	0.49	0.46	0.32	
Turkers	0.48	0.21	0.12	0.11	
Combined	0.35	0.31	0.17	0.07	

players would have been able to build on previously validated bridges and to remove bridges that they discovered not to work. Nevertheless, the negligible difference in average scores suggests that such feedback does not dramatically alter the system's assessment of causal understanding.

The scores reported here are significantly lower than those reported by (Plate, 2006, Table 6-9). While untrained university students in his study averaged 42.7 points on the rubric (using 2^{nd}-order calculations, the same used in *Causlings*), subjects in this study averaged only 10.6 points across both conditions—even the high score average, which can help offset multiple games by misunderstanding players, was only 20.9 points. However, as mentioned above, this game considered the influence direction of created causal relationships—players not only had to get the target direction

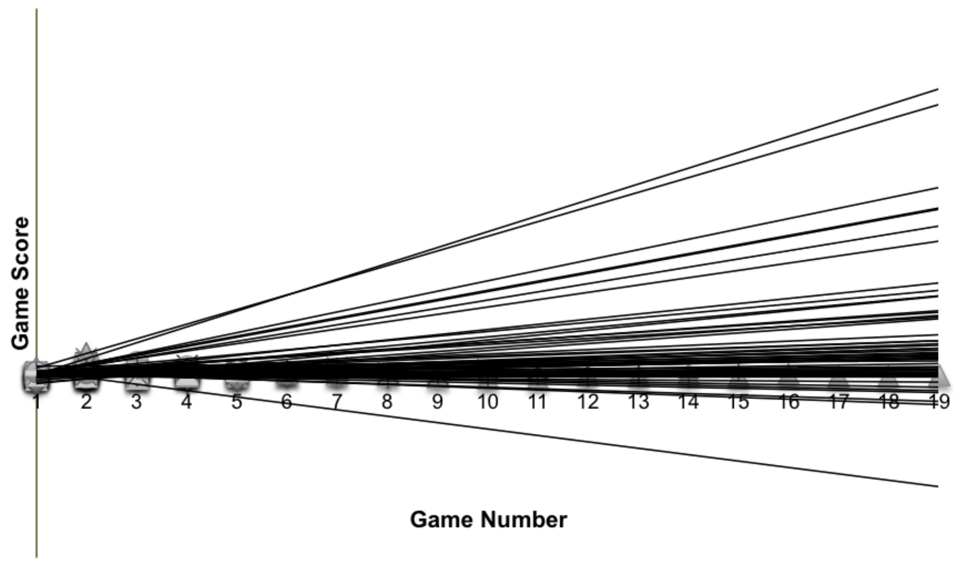


Figure 6.1: Trendlines for changes in scores over multiple games for each player.

correct, but also whether the relationship was increasing or decreasing. Thus it is understandable that players in this study received significantly lower scores[1]. Furthermore, *Causlings* scores may have been adversely affected by the change in study procedures from Plate's CMAST framework. For example, Plate had subjects explicitly group causal factors into categories, following the 3CM method (Kearney and Kaplan, 1997), before constructing a map—this procedure would allow subjects to gain familiarity with the concepts, as well as begin forming mental models of the components of the system. In my study, I forwent this procedure, as it did not directly related to assessing accuracy of causal maps in comparison to experts. Without such a procedure in my user study, players had to perform such island grouping on the fly while playing *Causlings* (and indeed, many did perform such categorization—see Section 6.2 and Section 6.3), making the task more difficult and potentially resulting in lower scores. Future research into increasing causal understanding might consider the benefits and uses of this simple categorization process.

[1]Differences in scoring metrics may also account for some discrepancies.

6.1.1 Correct and Incorrect Bridges

The game scores reported above are a direct function of which causal bridges players built while playing *Causlings*; considering whether frequently built bridges were correct or incorrect can further illuminate the effects of the interactive system on assessing causal understanding. Table 6.3 describes the breakdown of all the bridges built across all 333 games considered. Although more bridges were built in total in the basic mode condition (across fewer game; see Table 6.1), this result appears to be primarily due to a few outlier games that included a large number of bridges—on average, players in timed mode games built *more* bridges per game, though not significantly more. Watching the Causlings explore as they built their causal maps in timed mode may have encouraged players to build more bridges, whereas basic mode players were more likely to build a small map with only a few bridges, and then test the accuracy of that map before proceeding. This result does demonstrate that the time limits imposed by the game framing did not limit the size of the causal map that players would create. Furthermore, on average, games in the timed mode included a significantly greater number of correct bridges (though the same number of incorrect bridges) compared to in basic mode. This increase is possibly due to the feedback provided in the timed mode, which allowed players to remove incorrect bridges before the game ended. Another possibility is that because timed games finished faster and resulted in more games played, players had greater opportunities to learn (e.g., memorize) which bridges were correct and include them in later games. This result thus suggests that repeated playing can help people to learn to accurately identify causal relationships, though the question of transferring that knowledge to outside the game context remains unexplored.

Table 6.4 lists the 10 most frequently built causal bridges that were correct according to the expert maps used as a basis, ordered by frequency of appearance across all games (both modes). Note that this list also includes the top 6 correct bridges from each mode, though certain bridges were more common in one mode than the other. In both modes, the by far most commonly built bridge was "Samaki population increases Amount of samaki caught"—a bridge connecting the starting

Table 6.3: Number of causal bridges built by game mode.

	Basic Mode	Timed Mode	Both Modes
Correct bridges	1061	1079	2140
Incorrect bridges	1039	844	1883
Total bridges	2100	1923	4023
% correct bridges	50.50%	56.10%	53.20%
Avg. correct bridges per game*	5.8	7.2	6.4
Avg. incorrect bridges per game	5.7	5.6	5.6
Avg. total bridges per game	11.5 (median: 8)	12.7 (median: 11)	12.1 (median: 12)

** difference between game modes significant at p < 0.02*

causal factor ("Samaki population") with an island that was listed on the first page of options ("Amount of samaki caught"). As islands were listed in alphabetical order, this chosen island was in fact the second in the list from the left, putting it between the center islands (where the vision may unconsciously focus) and the beginning of the list (where English-speakers read from); in other words, a visually prominent position. Indeed, the 3^{rd}, 4^{th}, 5^{th}, and 7^{th} most common bridges also are all built between islands that are initially shown on the system's main screen, suggesting that the order of presentation and viewing may have a significant effect on the causal maps that are constructed and as such how player's causal understanding is assessed. Moreover, most of these bridges were more commonly built in timed mode than in basic mode, when players would have had less time to consider all the options for islands to include and bridges to build. These results further point to the effects of ordering and categorization on assessment, as well indicating potential difficulties caused by the complexity of the causal system being mapped.

As expected, frequently used correct bridges mirror causal relationships that are commonly considered. For example, "Amount of samaki caught increases Cost of samaki per unit catch" draws on the simple cause-and-effect relationship between economic supply and demand—a type of causal relationship that many people (and students in particular) are taught and understand. The bridges also reflect common everyday causal relationships, such as how increased effort leads to increased results with "Effort put into catching samaki increases Amount of samaki caught", as well as draw-

Table 6.4: Most frequently built correct causal bridges.

Bridge	# Basic Games	# Timed Games	Total Games	Rubric Score
Samaki population increases Amount of samaki caught	50	55	105	4
Effort put into catching samaki increases Amount of samaki caught	30	26	56	8
Amount of samaki caught increases Cost of samaki per unit catch	25	27	52	4
Amount of samaki caught decreases Samaki population	17	33	50	6
Samaki population increases Coastal water quality	29	19	48	4
Reproduction rate of samaki increases Samaki population	23	17	40	6
Bad weather decreases Amount of samaki caught	11	27	38	1
Public information to increase fish oil intake increases Demand for Omega-3 as a food supplement	22	12	34	6
Demand for Omega-3 as a food supplement increases Omega Corporation profits	15	16	31	6
El Nino increases Bad weather	12	19	31	1

ing on relationships explicitly described in the samaki article ("Bad weather decreases Amount of samaki caught" directly translates the article's claim that "low catches are due to bad weather"). Indeed, a number of subjects reported that they started building bridges with relationships that were intuitive—as one subject wrote: "I tried to make connections to the most obvious ones first" (P10, University student).

Note that although 8 out of 10 of these frequently correct bridges are increasing (positive) bridges, increasing bridges were the default built (as well as potentially being more intuitive) and accounted for 80.1% (3224 out of 4023) of bridges built across all games. In fact, one of the common misunderstandings of how the game was played involved confusion surrounding the meaning of increasing and decreasing bridges—players were unclear if the Causlings would still cross decreasing/negative bridges, or what exactly a negative bridge represented. I observed more than one player play a game with a large number of decreasing bridges, get a low score, and then forgo using decreasing bridges in the future: "I tried to do the negative once but I got a negative score so I didn't really mess with it" (P40, University student). Players would attempt to force their own meanings onto the influence direction of the bridges, as one player described: "If I thought it was a positive relationship where one thing could cause another thing, that would be the orange line [increasing bridge] from island 1 pointing to island 2. If it had an indirect effect I would use the negative from one thing to another. Just a guessing game I guess" (P30, University student). The frequency of this misunderstanding (itself probably due to flaws in the instructional presentation) likely had significant influence on which bridges were constructed and the scores players received.

Although bridges connecting initially shown islands may be more expected in the timed mode condition, it is unclear why "Amount of samaki caught decreases Samaki population"—the inverse relationship of the most frequent bridge, "Samaki population increases Amount of samaki caught"—was so much more common in timed mode than basic mode (being used almost twice as often). It is possible that in the timed mode players focused on making as many bridges between as few islands as possible, rather than considering other islands to use, and thus were more likely to make decreasing bridges between initial islands, except that timed mode games tended to include more islands (see Table 6.8). It is also unclear why "Samaki population increases Coastal water quality" and "Public information to increase fish oil intake increases Demand for Omega-3 as a food supplement" were used so much more frequently in games in the basic mode.

Note that Table 6.4 also lists the rubric values associated with each bridge (see Table 4.1 for details of these values). The table shows that the bridges built most frequently by players were not exactly the causal relationships most commonly identified by experts, though three or more experts identified 5 of the top 10 bridges created (earning a value of 6 or higher). Indeed, the number of experts who identified a causal relationship is significantly correlated with the number of games that included that bridge ($r = 0.209$, $p < .0001$). This result suggests that in aggregate, players produce maps that are at least correlated with those produced by experts in terms of correctness, even if the player-created maps are not representative of the expert maps. Players seem to produce causal maps that are more similar than not to expert maps.

Table 6.5 and Table 6.6 list the *incorrect* bridges most frequently built in the basic mode and timed mode conditions, respectively. Unlike with the correct bridges, there was not significant overlap between these lists, and as such they are presented separately (each table includes the appearance counts from each of the two modes and overall). Note that none of these incorrect bridges were built more frequently than the most common correct bridges—the distribution of incorrect bridges was wider than the distribution of correct options. Players made mistakes in a greater variety of ways than they made correct decisions, possibly due to there being a larger number of possible incorrect bridges to make.

Table 6.5: Incorrect causal bridges most frequently built in basic mode.

Bridge	# Basic Games	# Timed Games	Total Games	Correct at Degree
Samaki population increases Demand for Omega-3 as a food supplement	17	11	28	-
Samaki population increases Nutrients in the water	17	5	22	-
Samaki population increases Food eaten per fish (samaki)	14	4	18	3
Samaki population increases Reproduction rate of samaki	14	3	17	4
Amount of samaki caught increases Samaki population	12	8	20	3
Demand for Omega-3 as a food supplement increases Amount of samaki caught	12	19	31	3
Samaki population increases Lifespan of samaki	12	5	17	4
Samaki population increases Cost of samaki per unit catch	11	10	21	3
Samaki population increases Human population	11	9	20	-
Samaki population increases Management of samaki catch	11	4	15	3

In the basic game mode condition, overall the most frequently built incorrect bridges involved

Table 6.6: **Incorrect causal bridges most frequently built in timed mode.**

Bridge	# Basic Games	# Timed Games	Total Games	Correct at Degree
Demand for Omega-3 as a food supplement increases Amount of samaki caught	19	12	31	3
Demand for livestock feed increases Amount of samaki caught	13	3	16	3
Amount of samaki caught decreases Cost of samaki per unit catch	12	5	17	4
Demand for farm-raised fish feed (aquaculture) increases Amount of samaki caught	11	3	14	3
Human population increases Amount of samaki caught	11	2	13	4
Samaki population increases Algae Blooms/Dead Zones	11	10	21	4
Samaki population increases Demand for Omega-3 as a food supplement	11	17	28	-
Samaki population increases Cost of samaki per unit catch	10	11	21	3
Samaki population increases Human population	9	11	20	-

players specifying that the starting island "Samaki population" incorrectly increased another causal factor (e.g., "Demand for Omega-3," "Nutrients in the water", etc.). As the game instructed players to build causal maps starting from the "Samaki population" island, it is likely that players experimented (either thoughtfully or not) with bridges from this home island to others. Indeed, some of these incorrect bridges represent attempts to describe complex causal relationships using the islands provided—for example, "Samaki population increases Nutrients in the water" points at the role of the samaki in the ecosystem (the article states that the samaki provide nutrients to other fish), but according to the expert map increasing the samaki population does not *directly* increase the nutrients in the water. Other mistakes point to problems understanding the system dynamics of stock and flows (as described by Booth Sweeney and Sterman, 2000), such as believing that a larger fish population would directly increase "Food eaten per fish (samaki)" or the "Reproduction rate of samaki"—increasing the population would change the total stock of food consumed or babies produced, but would not change the *rate of change* per fish.

That said, both of these causal relationships are correct when considered indirectly from a "zoomed-out" view: while increasing the population does not directly influence these factors, it does have indirect influence and so is correct when considered using 3^{rd} or 4^{th} degree causality matrices respectively. Indeed, 7 out of 10 of the most common incorrect bridges represent problems identifying the *most direct* causal relationship, rather than problems identifying a causal relationship at all. It is possible that players either did not understand the indirection of these causal relationships, or that they were simple unaware that there was a more direct bridge that could be built due to not being able to consider all at once the large number of possible islands. Alternatively, players may have used incorrect bridges to try and connect otherwise correct portions of their constructed maps—for example, a number of players built maps around the "Human population" causal factor (see Section 6.2), and may have attempted to connect this factor to the home island of "Samaki population" using a default increasing bridge without thinking through the correct causal relationship.

For players in the timed mode condition, the most frequent incorrect bridges were those that reflected the indirect causal relationship between demand and production (i.e., demand for Omega-3, livestock feed, and fish feed leading to more fish being caught), but failed to trace the full path of the causality. Players would skip a step or two in constructing these bridges (such as "Effort put into catching samaki increases Amount of samaki caught," though this was one of the most common correct bridges), thus suggesting a failure to notice potential intermediate islands, or possibly a shallower understanding of the intricacy of the causal relationships in this complex system. Indeed, the five most frequent incorrect bridges in timed mode involved causal bridges pointed at the "Amount of samaki caught"—although players could often connect "Amount of samaki caught" to the home island of "Samaki population" (as one of the most frequent correct bridges), in the timed mode players had difficulty connecting this island to others. Again, this result is potentially due to the restrictions imposed by the game mode: players may have been unable to consider all of the islands and find the most direct connections in the time allotted.

Otherwise, players in the timed mode built similar incorrect bridges to those in the basic mode (such as mistaking the relationship between "Samaki population" and "Demand for Omega-3," or between "Samaki population" and "Human population"). There is some overlap between these two lists, with certain incorrect bridges common to both modes. These incorrect bridges suggest failures to understand the causal relationships that go beyond effects of the game mode, such as difficulty identifying the sub-steps of indirect causal relationships.

Overall, which correct and incorrect bridges players built seems to have been affected by the *Causlings* prototype's interface and presentation (i.e., what islands are shown on the initial page load, the presence of the timer and feedback from the dying Causlings, etc). But at the same time, this analysis, supported by my own observations, suggests that players struggled in particular with identifying the most direct causal bridges, rather than more abstract, indirect relationships. Building indirect bridges rather than direct bridges suggests that the players may be aware of causal relationships at an abstract level (e.g., how demand for by-products may influence fish stocks), but do not fully understand the intricacies and interrelationships that make up this complex causal network. Being able to identify the steps of the indirect link between causal factors is important for finding points of intervention that may allow people to effect changes in systems; and this interactive system prototype reveals some of the difficulties in considering these indirect links—particularly when the complex system in question has a large number of causal factors. I will further consider to what extent players include such interconnectivity and indirect paths in their causal maps in Section 6.2.

6.1.2 Supporting Learning

When asked what they learned from playing the *Causlings* prototype, nearly 40% of respondents described simply learning about the samaki and gaining a sense for why they should care about the fish[2]: "Before this game, I really had no idea what a Samaki fish was but after playing this game, I

[2]Remember that the samaki, although a fictional alias, was presented as a real fish.

was able to get a basic idea about what the fish looks like, some background information, and why the Samaki are important" (P30, University student). This knowledge-level learning (Bloom et al., 1956) was not enabled by playing the game, but rather by reading the introductory materials in a manner reminiscent of common educational practices—read the textbook, and then get tested on it. However, framing this content as the game may have increased motivation to read the otherwise "dull" (P48, Turker) article, such as how one player described: "I was focused when reading the long article, which I could never be if not because of this game" (P64, Turker). In this way, the game framing of the *Causlings* interactive system may have offered similar learning benefits to other "edutainment" systems (see Klopfer et al., 2009a). But as with edutainment games, *Causlings* did not support learning for everyone: almost 25% of players responded learning nothing from playing the game, possibly because they didn't see any benefit to knowing about the samaki, or due to frustration and confusion with the interactive system itself.

Nevertheless, a few players (around 15%) reported gaining deeper insights into the complexity of the environmental system's causal relationships from using the interactive system. One player reported: "It really got me thinking about causal relationships. It makes me realize that things are more complicated than they appear" (P60, Turker), potentially referring to how just having an overwhelming set of play options may have implied the scope of the causal system's complexity. Other players described learning about the non-linear relationships of the causal system: "I learned that a small effect on one thing can drastically change the whole setup. I had one bridge going the wrong way for a few seconds and seemed it put a plug in the whole thing and no Causlings could go anywhere. This showed me how a small change to something can domino into a much bigger problem" (P49, Turker). This player seems to have made a cognitive connection between the non-linear relationships in the complex system of the game (e.g., that certain bridges could act as bottlenecks for others), and the non-linear relationships in causal systems as a whole. Finally, as described above, some players performed the metacognitive learning task of reflecting on the accuracy of what they thought they knew: "I learned that there are many connections between a lot of things and sometime [sic] what you think is one is not a connection at all" (P4, High school student). In these ways, the interactive system prototype may have been able to support learning and systems thinking for a limited number of players—and again, adopting a more engaging topic than that used in this user study could further support such learning. On the other hand, the rarity of such feedback may reflect reporting biases, with players guessing at the purpose of the game and so reporting successful learning, whether such learning occurred or not.

6.2 What forms do the causal maps constructed by players take?

I begin considering the form given to constructed causal maps with what components (factors and relationships) were included in these maps. As described in Section 6.1, although there was a notable distinction in which incorrect bridges were most frequently used between game modes, the two modes had similar frequency distributions of correct bridges. Similarly, there were only small differences between the game modes in terms of which islands players chose to include in their causal maps. Table 6.7 details the frequency that each individual island was included across all games completed. Due to its centrality to the game, the most frequently used island was "Samaki population" (though seven games produced maps that did not connect this island to any others, likely due to misunderstanding how to play). The next most frequently used island was "Amount of samaki caught"—again, this island is an "obvious" causal factor that is easily visible as the game begins, and so is likely to be included. After these outliers, the other islands that appeared on the initial page are frequently used, along with "Effort put into catching samaki"—an island from the second page that was frequently linked to "Amount of samaki caught", as discussed above. Use of the remaining islands is loosely related to their alphabetical ordering, and otherwise follows a fairly regular "long tail" distribution. There appears to be no single pattern to these frequencies, and island use is likely a product of how intuitive or obvious the causal relationships are (e.g., the economic effects of the various "Demand for..." islands) or how important the causal factor is for mentally modeling the complex system (e.g., using "Human population" as a central component, as discussed below).

Table 6.7: Island inclusion frequency across all games, with degree of connection.

Island	# Games Included	# Basic Games	# Timed Games	Avg. Degree	Avg. # Tails (Out)	Avg. # Heads (In)
Samaki population	326	179	147	3.626	2.083	1.543
Amount of samaki caught	226	114	112	3.956	1.810	2.146
Demand for Omega-3 as a food supplement	147	75	72	2.435	1.333	1.102
Coastal water quality	144	71	73	2.431	1.056	1.375
Effort put into catching samaki	140	62	78	2.421	1.221	1.200
Cost of samaki per unit catch	137	63	74	2.540	0.927	1.613
Algae Blooms/Dead Zones	126	53	73	2.079	0.984	1.095
Bad weather	121	56	65	2.107	1.504	0.603
Lifespan of samaki	111	61	50	2.045	0.937	1.108
Management of samaki catch	106	57	49	2.085	1.094	0.991
Nutrients in the water	106	55	51	1.877	1.009	0.868
Amount of sport fish caught	103	41	62	2.301	0.854	1.447
Reproduction rate of samaki	101	65	36	1.950	1.079	0.871
Demand for farm-raised fish feed (aquaculture)	94	43	51	1.819	0.894	0.926
Demand for livestock feed	92	38	54	1.826	1.022	0.804
Disagreement over samaki population health	89	46	43	2.112	0.899	1.213
Human population	87	48	39	2.644	1.839	0.805
Management at the ecosystem level	79	43	36	2.241	1.051	1.190
Dissolved oxygen levels	79	34	45	1.696	0.823	0.873
Omega Corporation profits	79	45	34	1.987	0.506	1.481
Public information to increase fish oil intake	79	44	35	1.962	1.266	0.696
Public worry about decrease of samaki	76	49	27	2.132	1.026	1.105
Sales price per unit catch	72	38	34	2.417	0.806	1.611
Food eaten per fish (samaki)	72	47	25	2.014	0.833	1.181
El Nino	69	34	35	1.493	1.130	0.362
Scientific speculation of overfishing	67	38	29	2.104	1.224	0.881
Samaki industry leaders' claim of healthy fishery	65	32	33	1.862	0.938	0.923
Production from international fish oil competitors	63	38	25	1.714	0.857	0.857
Price of competing products (soybeans/vegetable oils)	61	38	23	1.984	1.033	0.951
Predatory bird populations	51	27	24	1.529	0.804	0.725
Sport fish populations	49	21	28	2.163	0.776	1.388
Sport fish health	43	23	20	2.093	0.814	1.279
Reproduction rate per unit fish	40	25	15	1.525	0.650	0.875
Soybeans sales	39	27	12	1.821	0.564	1.256
Marine mammals	34	13	21	1.706	0.765	0.941
Society affluence	31	23	8	1.871	1.000	0.871
All islands	**333**	**182**	**151**	**2.364**	**1.182**	**1.182**

Moreover, as detailed in this table, the vast majority of islands did not have any notable difference in frequency of usage between the basic and timed modes. However, a few islands—such as "Reproduction rate of samaki," "Public worry about decrease of samaki," and "Food eaten per fish (samaki)" were notably more common in the basic mode condition than the timed mode condition.

These islands may be easy to include but less important to the causal system as a whole, and so are treated as a kind of "second tier" of island to use that may have been excluded due to time restrictions in the timed mode. The time limit in the timed mode may have excluded certain peripheral islands as well as emphasizing the islands early in the alphabetical list.

Table 6.7 also details the degree of each island averaged across all the games. An island's degree is the total number of bridges connected to it—a statistic that can be further divided into the number of "tails" (bridges whose target direction points out from the island, making the island the tail of the bridge) and the number of "heads" (bridges whose target direction points in to the island, making the island the head of the bridge). Islands with higher degrees can be understood as more important or "central" to the causal map, as they are connected to more islands. Thus "Samaki population" has one of the highest average degrees—it is the most central island in the expert graph, and is in a way required to be centralized by the game's mechanic of having the Causlings begin there. However, the island with the highest average degree was "Amount of samaki caught" (with a higher number of heads than tails) indicating that many players identified the causal relationships that influenced the samaki catch, using that as an intermediate step in describing causal effects on the samaki population.

"Human population" and "Cost of samaki per unit catch" also had significantly higher than average degrees and thus were more central to the constructed graphs. I observed a number of players construct maps based around the human population, such as how one player reported: "At first I started from the population of samaki, and then after the third game I changed it to the human population and how it affects the things in the first place. And that decreased the mortality rate" (P11, University student). Considering the causal system in terms of humans' effects on it may have been an easier way for the players to mentally model the causal system (thinking about what effect they as humans had on the fish). Indeed, in addition to having high centrality, the use of the "Human population" island is significantly correlated (r = .146 p < .005) with game score, suggesting that complex causal systems may be more effectively understood when grounded in more easily relatable factors, such as everyday aspects of the subject's life. The high centrality of "Cost of samaki per unit catch" again points at the pervasiveness of framing complex environmental systems in terms of economics (see e.g., Dourish, 2010; Ross et al., 2010c). And the even higher correlation between including this island and game score ($r = .247$, $p < .0001$) suggests that people already have a decidedly accurate understanding of the causal interactions of market-based systems (at least at the micro-economic level).

Note that there were few differences in island degree between the timed and basic modes, many of which can be accounted for by the differences in island usage. For example, "Amount of samaki caught" had a higher average degree in the timed mode, but it was also used relatively more frequently in this condition and so likely had more bridges built connected to it (particular as during timed mode I observed players building extra bridges to move the waiting Causlings without giving careful attention to the bridges' meanings). So overall, there was not a significant difference between game modes in terms of which islands were used and how centrally they were connected. Again, these results suggest that the use of this particular game pattern in the interactive system may not have drastically influenced the causal maps constructed and the subsequent assessment of players' causal understanding—at least beyond limiting the time allowed to construct a full and complete map.

6.2.1 Graph Structure

By considering the graph representing each game's resulting map, I can consider some of the properties and overall trends in these maps. A selection of these properties is detailed in Table 6.8. On average, players used 10 or 11 of the 36 possible islands in each game they played (median 9)—with individual games ranging from only two islands connected by a single bridge[3] to complex maps of all 36 islands (see e.g., Figure 6.2). Indeed, although the implicit goal of the game was

[3]Two games used 0 islands—the player in timed mode twice ran out of time before she could build a bridge. As this player came from the Turker subject pool, it is possible that she loaded the game and then proceeded with a different HIT, letting the timer run out.

to use all the islands—"that's the challenge!" as one player put it (P19, University student)—the majority of games used less than 1/3 of the islands. As mentioned several times, this limitation is likely due to the overwhelming number of islands to choose from. Thus I interpret this result as continuing to demonstrate player's limited ability to map a complex causal system with a large variety of factors, with potential implications both for assessment techniques such as *Causlings* and for systems thinking pedagogies overall.

Table 6.8: Average properties of graphs of causal maps by game mode.

	Basic Mode	**Timed Mode**	**All Games**
# Islands Used	9.7	10.91	10.25
# Bridges Built	11.54	12.79	12.11
Link Density	1.07	1.11	1.09
WCI*	0.38	0.44	0.4
Diameter	4.18	4.56	4.35
Max Chain Length	4.12	4.03	4.08
Avg. Chain Length	2.08	2	2.04

** difference between modes significant at 0.05*

On average players built 12.11 bridges among their islands, producing an average link density of 1.09, or 1.09 bridges per island. Because each bridge is connected to two islands, this means that across all games, each island had on average 2.18 bridges connected to it—slightly more than if the island was part of a long chain. Again, this average spans across a range of link densities, with some causal maps including a dense web of connections (e.g., Figure 6.3), and some having a sparse tree of bridges (e.g., Figure 6.4 and Figure 6.5). In fact, in a number of games the player created just a single path of causal relationships, either because of confusion about the game[4] or simply because they thought through the causal connections linearly. Nevertheless, the vast majority of games included some level of branching, producing an average Web-like Causality Index (WCI) of 0.40—that is, 40% of islands had either more than one bridge pointing in or more than one bridge pointing out. Moreover, these measurements of the maps' interconnectivity, density,

[4]A few players tried to connect islands temporally, either based on their appearance in the article or based on which factors would occur before others.

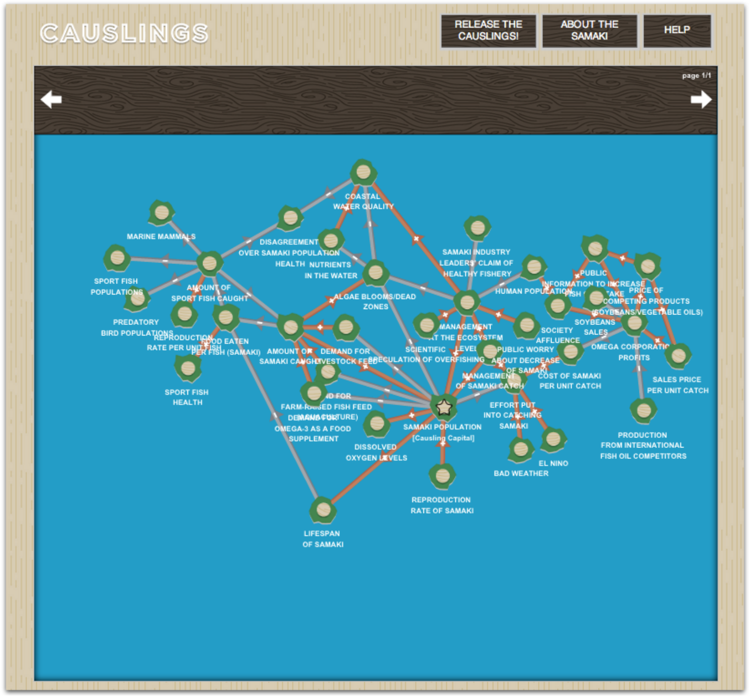


Figure 6.2: A player-built causal map using all 36 islands. This map has higher-than-average link density and WCI, and earned a rubric score of 22 points in basic mode.

and "web-like-ness" are comparable to the results of Plate's (Plate, 2006, Table 6-10) CMAST studies, particularly the maps produced by his university students before receiving systems thinking instruction (the group that would be most similar to the subjects of the study presented here). This similarity can be seen as potentially validating the cognitive mapping enabled through the *Causlings* prototype—while subjects may not be as accurate in the maps they produced without the initial categorization activity, the maps produced seem to have similar structure and interconnectivity; using the interactive system for assessment has not had a significant effect on whether subjects include web-like causality in their causal maps.

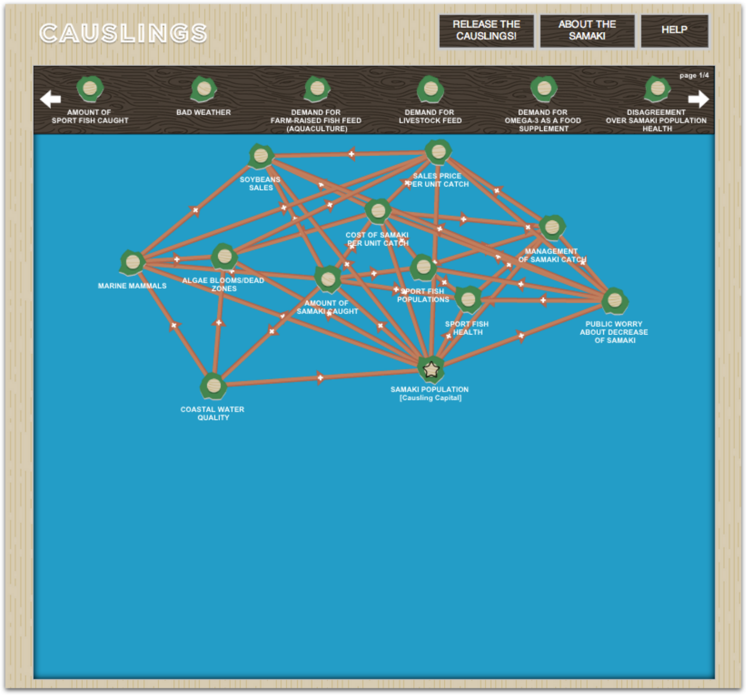


Figure 6.3: A causal map with very high link density (3.17) and WCI (1.42). This map earned a rubric score of -17 points, as many of the large number of increasing bridges were incorrect. Constructed in basic mode.

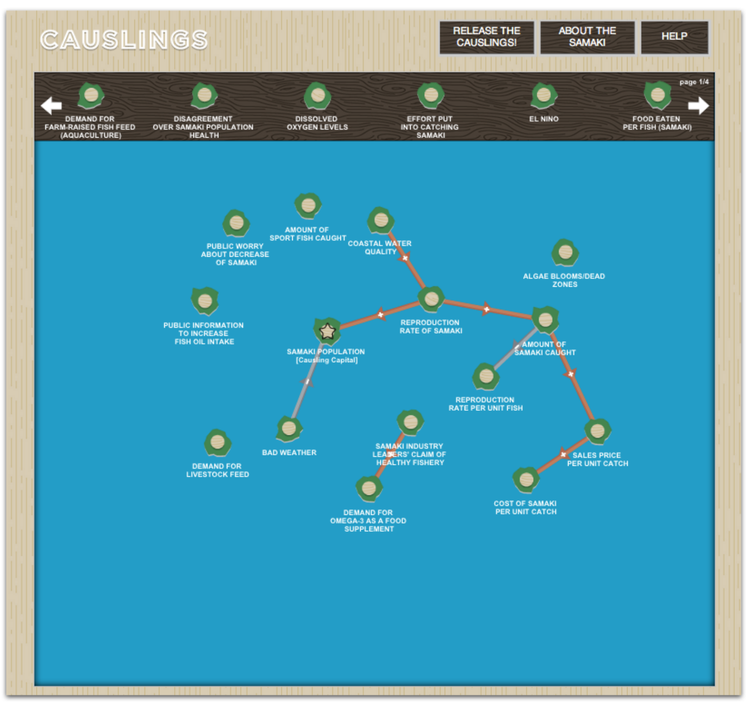


Figure 6.4: A causal map with relatively low link density (0.8) and WCI (0.3). This player in timed mode was not able to finish connecting islands, resulting in the low density. This map earned an overall rubric score of 1 point.

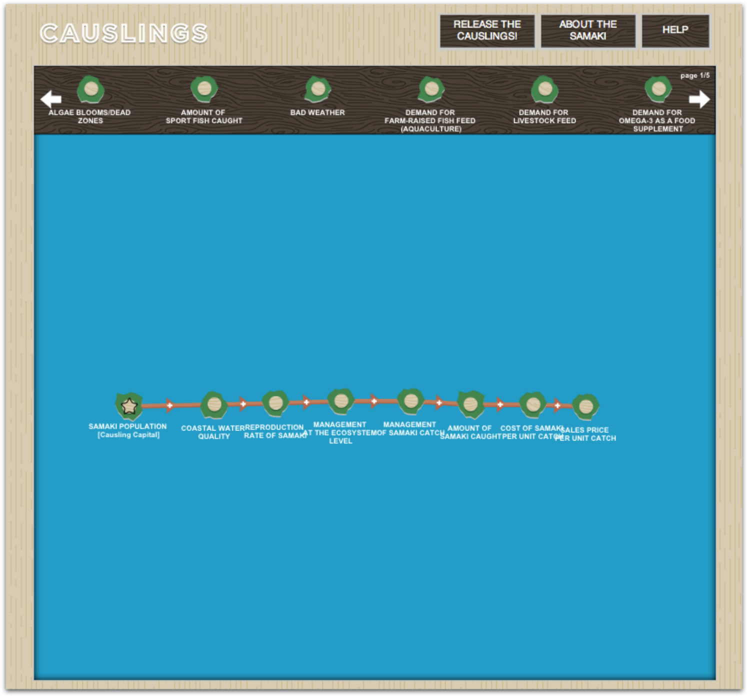


Figure 6.5: A causal map that is a single linear chain, producing a WCI of 0.0. This player produced this chain after 11 games of trials and continuous expansion (basic mode)—all the causal relationships are correct, earning an overall rubric score of 16 points.

Finally, Table 6.8 also details the average diameter of the causal maps (the maximum number of bridges, independent of direction, between any two islands), as well as the mean of each mode's maximum and average chain length—that is, the length of the longest and average directed path between islands, whether correct or not. These measurements can provide a sense of the overall "shape" of the graphs. An average diameter of 4.35 suggests that generally players did not build bridges more than three steps away from the central islands—if a map had "Samaki population" at the center, then a diameter of 4 would involve one level of indirect causal relationships. With an average maximum causal chain length of around 4 (as well as the multiple islands with high centrality), these statistics suggest that maps did tend to spread out, as illustrated in Figures 6.2, 6.3, and 6.4. On the other hand, an average causal chain length of around 2 suggests that that maps tended to be "short and fat," which Eden et al. (1992) suggest indicates "a high range of choice and alternative views" but "little depth of thinking". In other words, players built maps that attempted to include a wide variety of choices, but did not delve deeply into any particular causal relationship. However, this effect could be an artifact of the islands provided: players were given a "short and fat" system to model, and acted accordingly. In fact, the average chain length was *negatively* (though not significantly) correlated with game score—shorter chains produced better scores. This statistic further supports that the causal system considered focused on breadth rather than depth; even Plate suggested that the longest chain he could reasonably consider had length 4 (though the game detailed in Figure 6.5 demonstrated a correct causal chain of length 7). Thus the overall shape of the graphs is likely based on the shape of the causal system considered—the results of assessments of either more linear or more spread-out causal systems would likely be different.

Considered from a graph theoretical perspective, there was little difference between the causal maps produced in each of the two game modes. Surprisingly, players in the timed mode condition included slightly more islands and built more bridges on average than basic mode players, though neither of these differences was statistically significant. As such, this difference is likely a statistical artifact (possibly brought about by a large number of basic mode games that were ended "early" as part of player experimentation), though it may begin to indicate a higher level of engagement with the timed mode, such that players were willing to try including more islands in their causal maps—perhaps to better help the Causlings that were moving around. Similarly, graph diameter and average chain length were not significantly different. Yet while the link density was very similar between modes, the WCI of games in the timed mode condition was slightly significantly higher ($p < 0.05$). One possibility is that players in the timed mode may have been more likely to use more branching because of the feedback they received—watching Causlings fall of a bridge when leaving an island may encourage building more bridges from that island. Nevertheless, this difference is small enough that the graphs are still comparable across the two interaction modes.

Overall, the quantitative and statistical distinctions between the graphs produced in the two interactions modes have been shown to not be significant—maps produced in one mode were similar to maps produced in the other. Thus for the remainder of the analysis I consider the causal maps as a single set, particularly for analyzing their visual structure and their process of construction.

6.2.2 Organization and Visual Structure

Many players attempted to organize their islands and bridges as they constructed their causal maps, particularly through trying to categorize and group the islands. As one player explained: "I knew I was going to be able to group certain things together: certain economic factors needed to be with other economic factors" (P20, University student). Players would group the islands into categories around general topics (e.g., commercial factors, environmental and ecological factors, or agricultural factors) or around specific groups of islands (such as "Human population" or "the fish"). These categories overlap with those explicitly identified by subjects in the 3CM portion of Plate's (2006) CMAST study, suggesting that many players performed the same mental categorization step during the mapping process itself, even when not prompted ahead of time. In reifying their causal understanding, players needed to categorize different components of the causal system.

This categorization was often visible in the spatial layout of the causal maps, as shown in Figures 6.6, 6.7, and 6.8. Some players intentionally established this spatial organization, such as one who explained: "I had groups that all connected to the center. So if it was about cost management and sales of the fish then it would be in one area, and another area would be nutrition of the water and the quality of the coastal water and then the lifespan of the fish, and then another place would be for the concern of Omega-3, fatty acids, and things like that" (P23, University student). The creator of Figure 6.7 even tried to organize the direction of his bridges: "I almost tried to make it like layers going left then right in order, and then maybe doing different subjects. I was just playing with different things. That's what happens when you have an open playing field. You just move things around and fit them in" (P19, University student). But for many other players, this organization seemed to emerge organically as they tried to fit all the islands and bridges they wanted into their maps. Although Figure 6.8 is spatially organized, its creator claimed: "Really I just made it up as I went. You could see this affects that, but then you could see that there are some intermediate pieces that have to go between them, some that actually affect that, so it's more of a build and edit as you go" (P16, University student). Nevertheless, use of spatial metaphors such as "go between" points to the player trying to keep the map organized, even if it was not through initial categorization. Indeed, organized maps may have simply been a response to the complexity of the cognitive mapping process or even the interface itself; one player reported: "I was just trying to categorize it. Anything that was more specific to a particular section, I would just take it closer to that, just so I wouldn't have to make the bridges so long" (P34, University student).

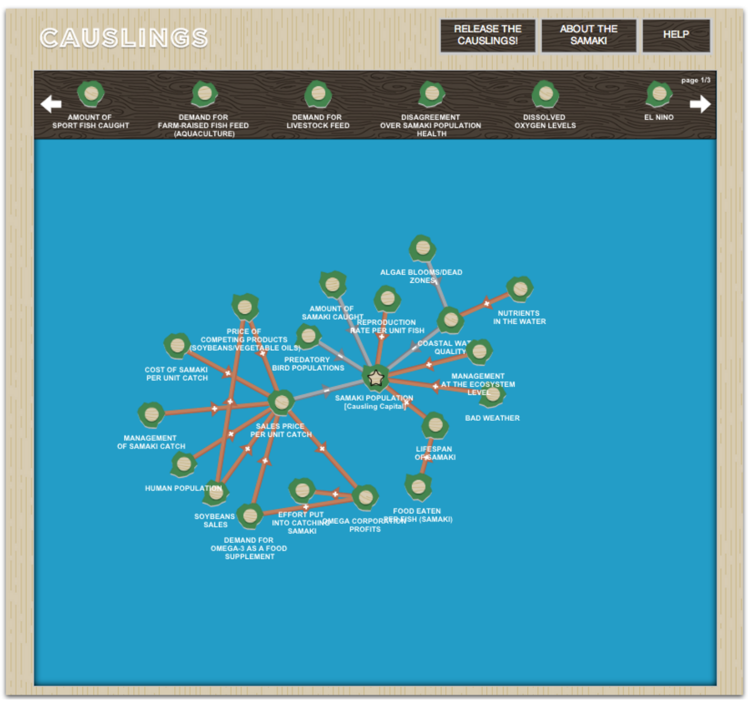


Figure 6.6: A causal map with clearly grouped islands. One group surrounds "Samaki population" and deals with ecological issues, and the other group surrounds "Sales price per unit catch" and deals with economic issues. Map constructed in timed mode, and earned a rubric score of 65 points.

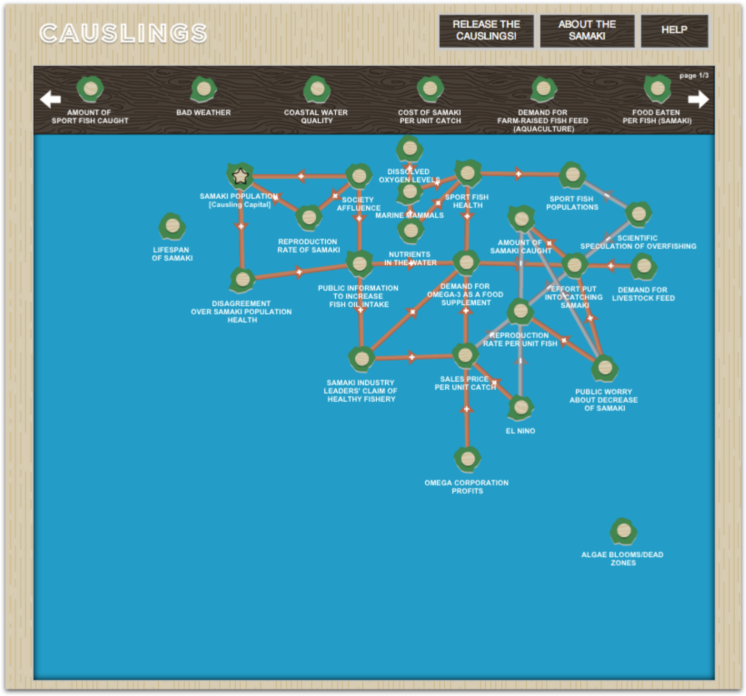


Figure 6.7: A causal map that is spatially organized (topics positioned from left to right). Only a few players laid out their causal maps in clean grids—and even maps that started on a grid often become more chaotic over time. Map constructed in basic mode, and earned a rubric score of 8 points.

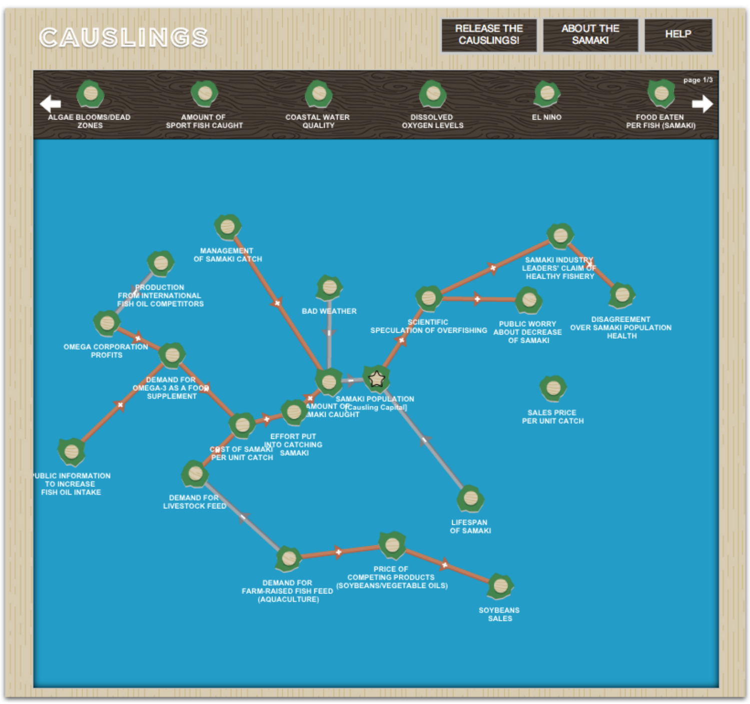


Figure 6.8: A causal map coincidentally organized by category. Each branch of the map consists of a different topic: e.g., the branch on the far left deals with fish oil and Omega-3, while the branch on the right deals with the public controversy over the samaki. Map constructed in basic mode, and earned a rubric score of 29 points.

Although many players attempted to categorically organize their causal maps, the time restrictions in the timed game mode hindered such efforts. For example, Figure 6.9 shows a causal map in which the islands seem to be organized (the islands are organized in a causal chain relating the "Samaki population" to the "Scientific speculation of overfishing," and issues of environmental management are grouped in the top right), yet the player ran out of time before she could finish building bridges between the islands. The time pressure in the timed game mode may also have encouraged players to skip carefully organizing islands. Instead, players would position islands almost at random, connecting them as needed and resulting in complex, haphazard maps such as that shown in Figure 6.10. Nevertheless, even in basic mode many players did not perform any conscious organization, instead "just putting whatever would fit" (P27, University student) in order to construct a map. Overall, maps that were more consciously organized earned higher scores (though the highest scoring maps had less explicit organization)—a relationship reflected on by some players, such as one that said: "I was way more organized on my last round and that was obviously the one I got the most points in" (P37, University student). In this way, these results from *Causlings* as an assessment tool continue to emphasize the importance of mentally categorizing aspects of causal systems in understanding those systems.

It's worth noting that in the majority of games players avoided having bridges cross one another (as most prominently demonstrated in Figures 6.3 and 6.10)—and even maps with overlaps would often have only a single crossing bridge. My data do not support whether players avoided these crossing consciously (e.g., in trying to keep the map cleaner, organized, or more readable), or whether the low number of crossings is an artifact of the cognitive mapping processes—for example, if players tended to consider causal systems in a planar fashion, rather than with such interconnectivity that would require bridge crossings. Yet at the same time, a significant minority of players did cross bridges in their causal maps. The effects of planar organization on the accuracy of causal maps—as well as whether such organization more or less frequently occurs in paper-based cognitive mapping systems—remains a question for future research.

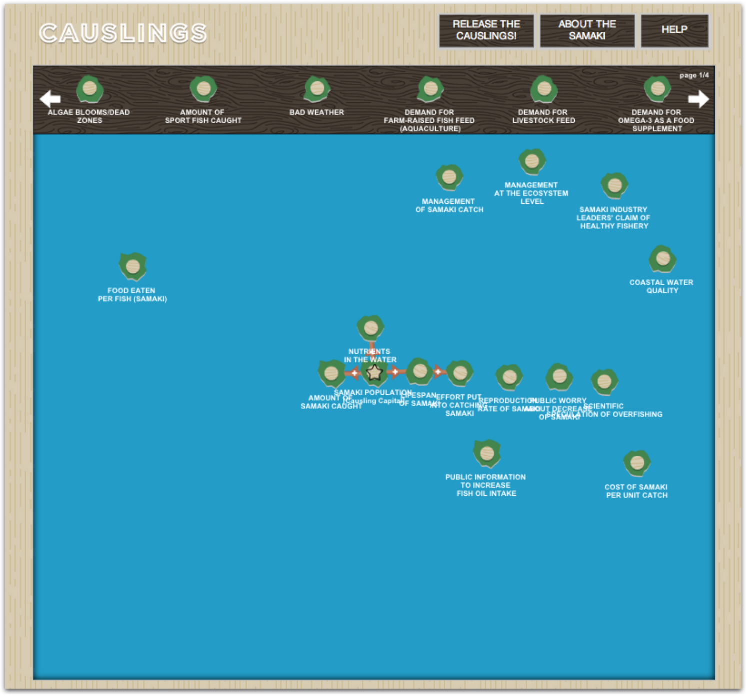


Figure 6.9: An organized causal map that was not completed due to time limits in the timed mode. Map earned a rubric score of -2 points (penalized for incorrect bridges).

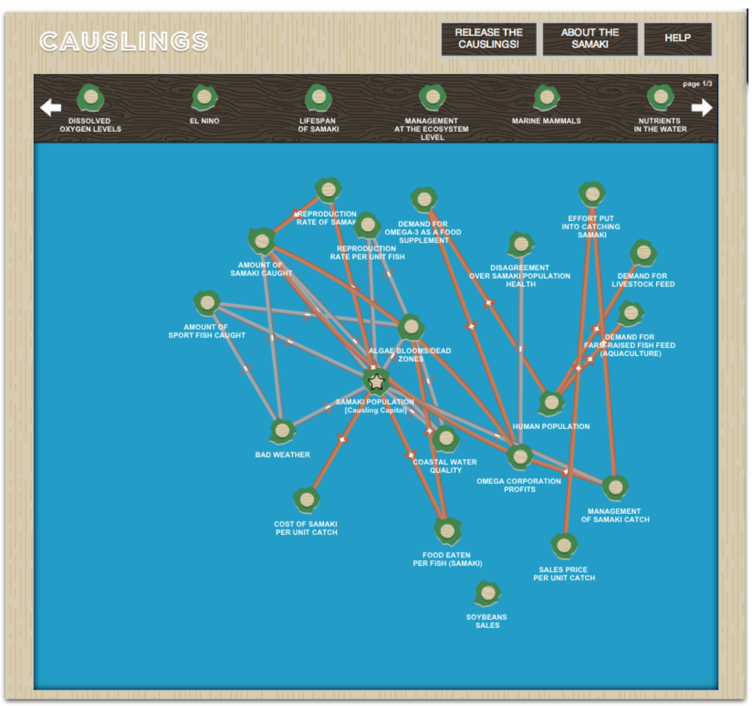


Figure 6.10: An apparently unorganized causal map. Constructed in timed mode, and earned a rubric score of 13 points.

The planar, tree-graph like branching seen in Figure 6.8 (and to a lesser extent in Figure 6.6) was a fairly common pattern in the constructed causal maps. Indeed, many graphs—particularly earlier, lower-scoring games—involved the creation of clear star-like patterns, such as in Figure 6.11. Often (but not always) building off the Causling Capital of "Samaki population," these graphs likely represent players focusing on the game framing's objective of helping the Causlings spread out, rather than building a causal map that would coincidentally allow the Causlings to explore. Other maps, such as in Figure 6.12 (as well as in Figures 6.2, 6.3, 6.7, etc.), had a greater numbers of graph cycles and islands with high degree, revealing a more intricate understanding of the causal system. Overall, maps with higher connectivity often had higher scores, if only because they included a greater number of bridges and so could score higher. Figure 6.13 shows the highest scoring map from my user study; this map is highly interconnected, and yet still has a clear "center" island in the "Samaki population." Whether giving certain islands high levels of centrality is intentional or an emergent property of the cognitive mapping process will be discussed further in the next section.

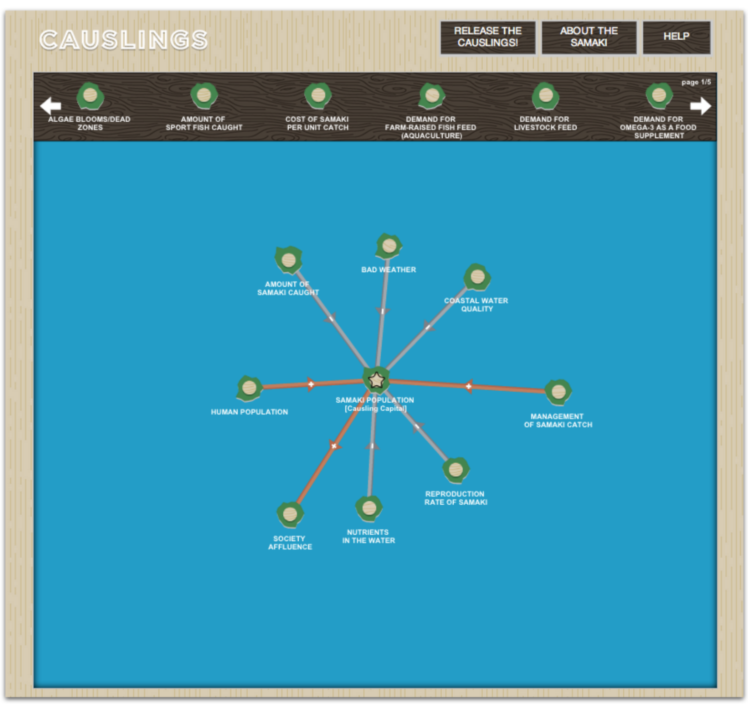


Figure 6.11: A star-shaped causal map. Constructed in basic mode, and earned a rubric score of 17 points.

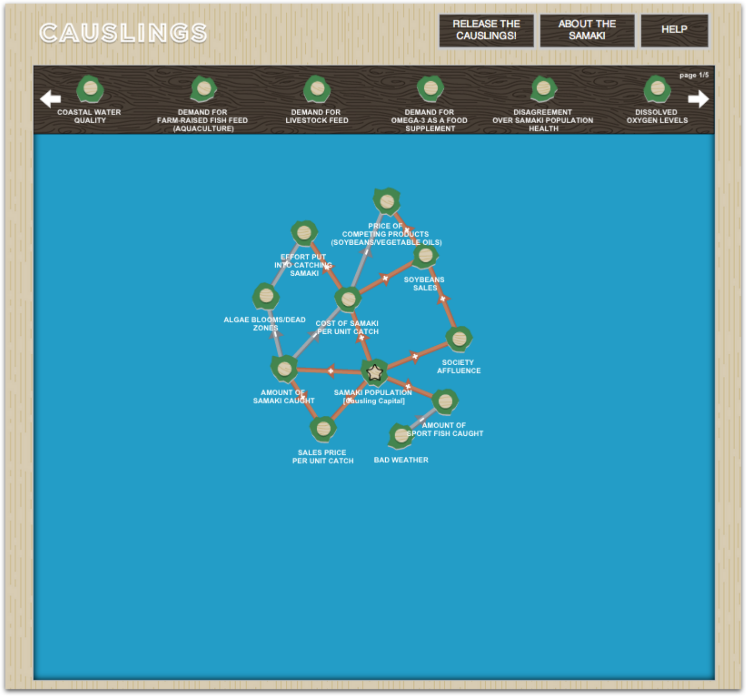


Figure 6.12: A densely connected web-like causal map. Constructed in basic mode and earned a rubric score of -10 points.

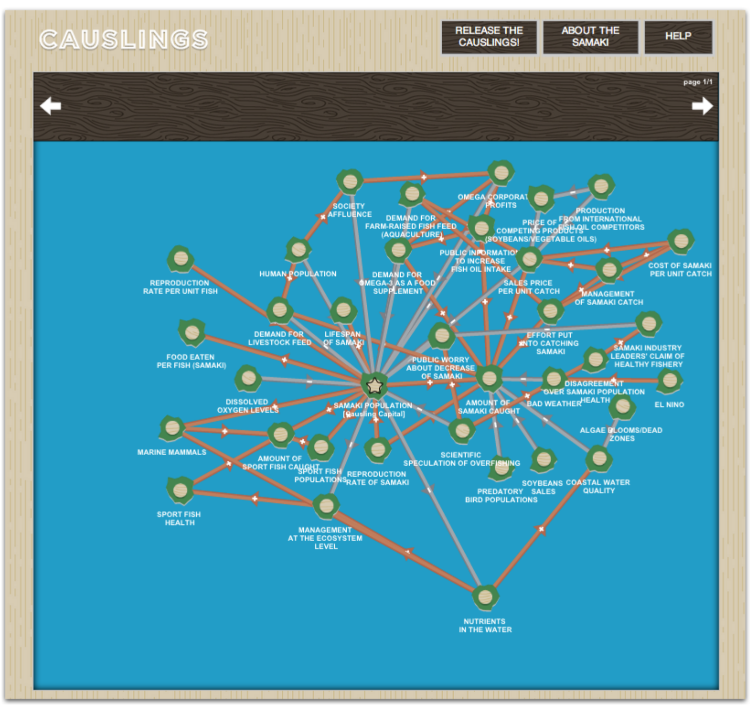


Figure 6.13: The highest scoring causal map from the user study, earning a game score of 76 points. Constructed in basic mode.

6.3 What patterns can be found in players' sequence of map construction?

In this section, I describe my explorations of the process and sequence through which players constructed causal maps while playing the *Causlings* prototype. As there was little to no significant difference between maps constructed in the basic and timed modes, I consider the building sequence across all causal maps as a single set. And as described in Chapter 5, this exploration is divided into three specific sub-questions, each of which is addressed in turn in its own subsection.

6.3.1 Do players select islands or create bridges in a particular order?

Table 6.9 lists the median and mean order in which islands were used—that is, the typical position in the sequence of islands used (where 1 is the first island used in a game, 2 is the second island, etc.). I consider an island first "used" when it is included as a factor in a successfully built bridge—in creating a bridge, the cause or "tail" factor of the relationship is considered to be used before the effect or "head" factor. Islands could also be interacted with (i.e., moved around the screen) before being connected with a causal bridge. Indeed, many players would drag a large number or even all of the islands onto the map area before beginning to build bridges. However, a player moving an island did not ensure that the island was included in the causal map at any point; as such, I denote islands as "used" when they have been connected as part of the map. Nevertheless, Table 6.9 also includes the median "touch" order, or an island's position in the sequence of islands interacted with (either moved or connected).

The islands players used earliest were also those that were most commonly included—the early alphabetical islands on the first page when the game initially loads. Moreover, the Causling Capital island "Samaki population" was significantly more likely to be the first island used; because I have usage order follow the target direction of each causal bridge, this means that players were most

Table 6.9: Median and mean order of island usage.

Island	Median Order Used	Mean Order Used	Median Order Touched
Samaki population	1	1.73 (s.d. 2.32)	2
Amount of samaki caught	3	3.96 (s.d. 3.68)	2
Coastal water quality	5	5.49 (s.d. 4.01)	5
Bad weather	4	5.96 (s.d. 4.78)	4
Amount of sport fish caught	4	6.09 (s.d. 5.06)	3
Algae Blooms/Dead Zones	5	6.15 (s.d. 5.31)	4
Cost of samaki per unit catch	6	6.72 (s.d. 4.67)	6
Demand for Omega-3 as a food supplement	7	7.65 (s.d. 4.46)	7
Effort put into catching samaki	8	7.72 (s.d. 4.21)	8
Demand for farm-raised fish feed (aquaculture)	7	8.37 (s.d. 5.08)	8
Human population	8	8.44 (s.d. 5.06)	10
Reproduction rate of samaki	7	8.48 (s.d. 5.99)	9
Lifespan of samaki	7	8.56 (s.d. 5.52)	9
Demand for livestock feed	8	8.59 (s.d. 4.78)	8
Public information to increase fish oil intake	9	9.30 (s.d. 6.16)	9
Management at the ecosystem level	7	9.35 (s.d. 6.34)	9
Management of samaki catch	9	9.41 (s.d. 6.11)	9
El Nino	9	9.83 (s.d. 7.14)	11
Food eaten per fish (samaki)	9	9.84 (s.d. 6.29)	10
Scientific speculation of overfishing	8	9.85 (s.d. 5.91)	10
Predatory bird populations	8	10.00 (s.d. 7.02)	12
Nutrients in the water	8	10.00 (s.d. 6.80)	11
Disagreement over samaki population health	10	10.09 (s.d. 6.14)	10
Samaki industry leaders' claim of healthy fishery	9	10.16 (s.d. 6.84)	10
Dissolved oxygen levels	8	10.33 (s.d. 7.47)	9
Public worry about decrease of samaki	9	10.35 (s.d. 6.02)	10
Marine mammals	8	10.49 (s.d. 7.67)	12
Sales price per unit catch	9	10.49 (s.d. 5.88)	11
Sport fish populations	9	10.82 (s.d. 8.60)	11
Production from international fish oil competitors	11	11.37 (s.d. 6.64)	13
Omega Corporation profits	12	12.06 (s.d. 6.02)	13
Society affluence	10	12.09 (s.d. 7.69)	11
Sport fish health	10	12.67 (s.d. 8.36)	11
Price of competing products (soybeans/vegetable oils)	13	12.84 (s.d. 5.30)	14
Reproduction rate per unit fish	14	13.41 (s.d. 7.60)	14
Soybeans sales	15	15.33 (s.d. 7.20)	16

likely to build their first bridge *from* the capital *to* another island (usually "Amount of samaki caught"). This result indicates how players may have thought—at least initially—of building bridges "outward" from the capital, following the narrative framing established by the *Causlings* game. After these two initial islands, there is a steady but overlapping trend in which order the islands were used—few islands stand out as overwhelmingly being used in a certain order. In fact, the standard deviations of island usage order are very high, indicating the wide spread of the ordering. Moreover, because each causal map only used about 10 islands on average (see Table 6.8), islands with median positions greater than that were used in less the half the games played; these islands may have been used later on average simply because they were mostly used in games that included more islands. Because games used different numbers of islands, ordering numbers may cause certain islands to seem to appear early—a particular island may be the last one used in a game, but it was still the fourth island used because the game only included four islands.

As a specific point of exploration, the island that was significantly most likely to be used the latest was "Soybean sales." Additionally, the keyword related island "Price of competing products (soybeans/vegetable oils)" was most often used later than almost all the other islands (though on average earlier than "Soybean sales")—indeed, it has a lower standard deviation than other late-usage islands, suggesting that it was more consistently used late in the game. Interestingly, during interviews players remarked upon these particular islands during a specific context: returning to and reviewing the article about the samaki[5]. One player explained that he reviewed the article looking for specific details: "One of the islands mentioned soybeans, so I was like 'hmm, where did it mention soybeans?'" (P16, University student). Another player described a similar process, pointing at the topic of soybeans as a specific impetus to refine her understanding:

> "The more I went on the more I remember seeing soybean. And I couldn't remember why soybean was there, so I was like it would really make sense for me to go back and figure this out. So I went back and then I looked over soybean, I looked at the effects it had on fishing, I looked at how other fish were becoming skinnier, how overfishing was causing problems for a lot of other stuff, how it decreased soybean sales and things like that. And so I think that was what made me more confident later on [to include those islands]... after a while I was like 'okay I can't keep doing the same four things, because my score is not going to get better'" (P35, University student).

In this way, the island of "Soybean sales" acted as a focal point for how *Causlings*' game-based framing (i.e., the desire to improve a score) may have driven player learning and even increased

[5]Only in about 13% of games did players look back at the article for more than 5 seconds at a time.

the player's self-confidence in being able to construct a causal map using the system.

Moreover, the fact that players felt they needed to review the article to remember the causal relationships involving soybeans and that soybean-based islands were generally used late in the game suggests that players did not readily retain information about this topic during their initial reading of the article. Yet "Soybeans sales" is a rather specific factor, even involving a unique keyword ("soybean") that jumped out at players for future review. While it is possible that soybeans were just not interesting enough to hold players' focus, it is also possible that players were better able to understand more general causal relationships (e.g., the relationship between "Amount of samaki caught" and "Samaki population") than relationships involving more concrete factors. Players seem to have more readily included abstract factors than concrete factors in their causal maps, potentially suggesting that it may be more difficult for players to construct causal maps and explicate causal relationships among specific, concrete factors—such as those involved in their everyday lives.

Because of the large number of possible bridges that could be built, it is not feasible to consider the median order of bridges built in the same way as I consider the order of island usage. However, I do consider how often pairs of bridges show up in sequence—that is, how commonly do players build a particular bridge X immediately before building another bridge Y. The most common bridge pairs are listed in Table 6.10. Bridge ordering was rarely identical—even the most common pair sequence was shared by only 14 out of 333 games (4.2%). Even when considering for bridge pairs with any number of bridges between them (that is, bridge X was built at any time before bridge Y; see Table 6.11), each pair's ordering was shared by less than 10% of games.

With both cases, the most common pair of relationships involved the most commonly created bridge (between the two most frequently and earliest used islands) preceding the third most common bridge. Indeed, these two bridges share a common island, and this repeated pattern involved the player creating an increasing causal path from "Samaki population" to "Amount of samaki caught" to "Cost of samaki per unit catch." Furthermore, in 10 of the 14 games in which this pat-

Table 6.10: Most common ordered bridge pairs (with no separation). The next most common pairs appeared in only 6 of 333 games.

1^{st} Bridge in Pair	2^{nd} Bridge in Pair	Games
Samaki population increases Amount of samaki caught	Amount of samaki caught increases Cost of samaki per unit catch	14
Samaki population increases Amount of samaki caught	Samaki population increases Amount of sport fish caught	10
Samaki population increases Effort put into catching samaki	Effort put into catching samaki increases Amount of samaki caught	9
Demand for Omega-3 as a food supplement increases Effort put into catching samaki	Effort put into catching samaki increases Amount of samaki caught	8

Table 6.11: Most common ordered bridge pairs (with any amount of separation). The next most common pairs appeared in 17 of 333 games.

1^{st} Bridge in Pair	2^{nd} Bridge in Pair	Games
Samaki population increases Amount of samaki caught	Amount of samaki caught increases Cost of samaki per unit catch	31
Amount of samaki caught decreases Samaki population	Effort put into catching samaki increases Amount of samaki caught	22
Samaki population increases Amount of samaki caught	Effort put into catching samaki increases Amount of samaki caught	19
Samaki population increases Amount of samaki caught	El Nino increases Bad weather	18

tern appeared, it represents the first two bridges created. However, 6 of the 14 of those games earned overall negative scores; the average score among games with this pattern was 4.71, significantly lower than the rest of the games on average. It is unclear what may have caused multiple players to share this building sequence. It is possible that the first bridge was constructed as the "obvious" choice (as described above), and that the second bridge was simply the players (mis)-identifying a connection between the supply of fish caught and the resulting price of fish; the players attempted to consider the causal system from an economic standpoint but made mistakes in this interpretation. Indeed, the fact that the third island in this sequence was on average the sixth island used overall (rather than the third) suggests that although this was a shared pattern, it is not representative of the players as a whole, and may instead indicate a shared misunderstanding of either the interactive system (e.g., what an increasing bridge represents) or of economic causality (e.g., the effect of increased supply on price).

Overall, there were few patterns to the sequence of bridges built shared among a significant number of players. This potentially indicates an individuality of cognitive mapping; each player reifies the causal system in their own way, with few shared specific techniques. But in short, this exploration reveals no clear pattern in the sequence of adding causal relationships to a causal map, possibly because no such pattern exists.

6.3.2 Do players construct causal maps hierarchically?

In order to explore whether players construct causal maps following any sort of hierarchy, I look at to what extent the order in which players build bridges matches either a depth-first or a breadth-first traversal of the causal graph. I measure these traversals using three metrics: (1) the expansion to new islands, (2) the connectivity of new bridges, and (3) the sequence of change in distance from the first bridge built. The summative results of these metrics are presented in Table 6.12. For all of these results, I consider only maps that had at least 3 bridges built and so can have a distinct order—a set of 302 games.

Table 6.12: Average tendency towards depth-first or breadth-first building in the construction of causal maps. Numbers are the fraction of bridges that are either built "depth-first" or "breadth-first." Depth-first "distance from first bridge" is the average length of a depth-first chain (see below for details).

	Mean (all games)	Median	Corr. w/ Game Score
Expanding to new islands			
Depth-first	0.32 (s.d. 0.19)	0.27	-0.350***
Breadth-first	0.27 (s.d. 0.14)	0.25	0.097*
Building new bridges			
Depth-first	0.54 (s.d. 0.15)	0.50	-0.285***
Breadth-first	0.48 (s.d. 0.12)	0.48	-0.023
Distance from first bridge			
Depth-first	1.29 (s.d. 0.99)	1.10	-0.153**
Breadth-first	0.78 (s.d. 0.23)	0.85	-0.021

* $p < .05$ ** $p < .005$ *** $p < .0001$

First, I consider how the causal map expands to include new islands. For each new island added to a graph, I counted whether the island was connected via a causal bridge to the most recently connected island (irrespective of the bridge's target direction)—that is, whether the bridges connecting a new island built on the previous bridge connecting a new island. This count measured how "depth-first" islands were added: how likely they were to be connected along the same lengthening path of bridges. I also counted whether new islands were connected to the same island as the preceding new island (e.g., if Island 3 was added to the graph by being connected to Island 1, immediately after Island 2 had been added to the graph by being connected to Island 1). This pattern represents a "breadth-first" approach to adding new islands to the graph, with new islands growing breadth-first off a single core island. Table 6.12 lists the percentage of islands that were considered added either depth-first or breadth-first.

Based on these counts, *Causlings* players were significantly more likely ($p < .005$) to add new islands in a depth-first manner than in a breath-first manner—possibly because in sorting through the large number of available islands, players were most focused on the most recently added island when grabbing another to insert into the map. However, adding islands depth-first was negatively and significantly correlated with game score, while breadth-first addition was positively correlated (though much less significantly) with game score. Players who added new islands by connecting them to other recently added islands were much more likely to earn lower scores in the game, while players who connected groups of new islands to a single location were slightly more likely to get higher scores. One possible explanation is that adding islands depth-first tends to produce causal chains (Island 1 is connected to Island 2 which was connected to Island 3, etc.; the map in Figure 6.5 has a high percentage of depth-first island additions), with higher amounts of depth-first adding often corresponding to longer chains. Breadth-first island adding, on the other hand, was more likely to produce star-like maps (as in Figure 6.11), because of how new islands are connected to a single root. It is likely that players who added islands in a depth-first manner were more likely to be thinking about the graph in terms of causal chains—even a single causal chain—and thus were less accurate in mapping the entire causal web. Breadth-first adding, on the other hand, helped the maps to "spread out" and include multiple connections between disparate categories. Therefore these results suggest that players who built maps broadly rather than deeply and avoided long causal chains had higher levels of causal understanding; whether the building strategy leads to that understanding or the understanding directed the building strategy is unknown.

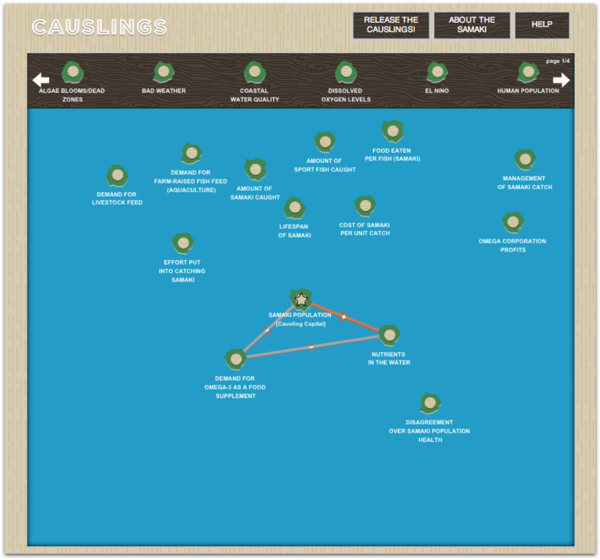
Second, I performed a similar count of depth-first and breadth-first additions for all bridges built, whether between new islands or previously connected islands. With this measurement, I considered a bridge to be built depth-first if it was connected to the alternate terminus of the bridge built immediately prior (so once a bridge from A to B was connected to Island A, the next bridge would need to be connected to Island B, and so on). Bridges were built breadth-first if they shared the same terminus as the previously built bridges. Again, Table 6.12 lists the percentage of bridges that were considered built either depth-first or breadth-first. As with adding islands, players were more likely to add bridges in a depth-first manner than in a breadth-first manner, and depth-first bridge building had a significant negative correlation with score. This suggests that players would build bridges in depth-first chains even between already included islands—again, such linear thinking suggests a less accurate understanding of the causal relationships in this complex system (note that for many smaller maps, each new bridge also connected a new island, making the these two measurements identical and supporting the matching correlation). However, there was not a significant correlation between breadth-first bridge building and game score; as the quintessential breadth-first bridge building pattern would be connecting a single island to all the others as appropriate, this results suggests that such an algorithmic consideration did not necessarily elicit accurate causal maps. Players who regularly made disparate connections revealed a greater understanding of the causal system—an understanding that potentially supported their ability to not use depth-first or breadth-first patterns in their map construction.

Third, I consider whether the map as a whole grew in a depth-first or breadth-first manner by considering the graph distance of each bridge from the first bridge built[6]. I measure distance to a newly added bridge as the length of the shortest path between the islands of the added bridge and

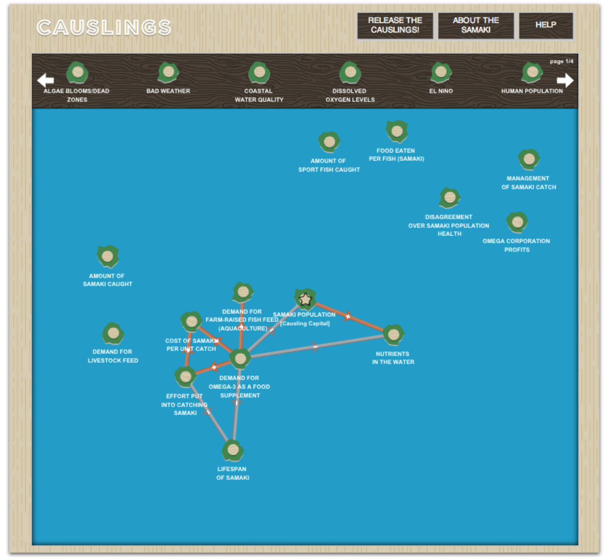
[6] I rejected measuring distance to the Causling Capital, as this island was not necessarily the center of each map. Indeed, the "center" island of each map could not be readily and consistently calculated for many well-balanced maps. Considering distance to the first bridge built best reflects my exploration of the growth of the causal map over time.

the islands of the first bridge built. Path length is calculated based on the matrix of the completed map—although this means that some bridges may move closer or further away as the map is built, it allows me to consider how the various pieces of the map grow (e.g., if the player builds a "far away" section of the map first before connecting it in such a way as to change its distance from the first bridge). For each bridge added, I calculated the distance from the first bridge built to produce a sequence of build distances (bridges that were disconnected from the first bridge built—either by being deleted or because the graph remained disconnected—were ignored). I measured how depth-first this sequence was by calculating the average sequence length of strictly increasing distances—that is, the average length of increasing distance represents the player's tendency to build depth-first. A player's tendency to build breadth-first was calculated as the percentage of bridges that were built the same distance from the first bridge as the previously built bridge. While not directly comparable, these two measurements allow me to consider the overall extent that a particular map was built either depth-first or breadth-first.

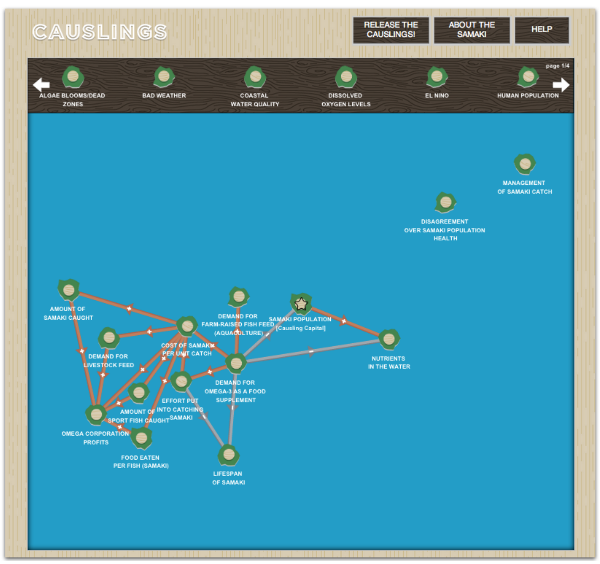
Overall players did not build causal maps that were particular deep (remember that the average graph diameter was only 4.35)—indeed, more than 1/3 of games did not include any bridges built that were more than 1 step away from the first bridge. Bridges were very rarely built in a depth-first sequence of average length greater than 2, indicating that few players expanded the entire graph in a depth-first manner. And depth-based building was negatively and significantly correlated with game score (the same low-scoring games with high depth-based measures in the previous metrics have high depth-based measures in this metric). On the other hand, considered with this metric, players were much more likely to build their graphs in a breadth-first manner—more than 78% of bridges were built at the same distance from the start as the previous bridge. Figure 6.14 provides an example of such building—the player would connect a set of islands to the graph in a breadth-first manner (one step away from start), then choose one of those islands to connect more to (now two steps away from start), etc. Anecdotally, this style of bridge building was quite common, as players would discover new islands to use and try to add them to the already defined graph without changing what had already been established. One player described this unwillingness to



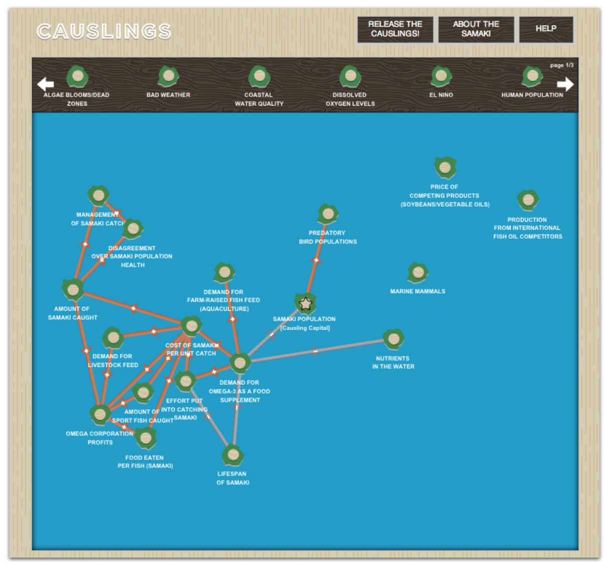
(a) At 68 seconds



(b) At 120 seconds



(c) At 168 seconds



(d) At 252 seconds (end)

Figure 6.14: A map being built in a breadth-first manner. Constructed in timed mode, and earned a rubric score of 15 points.

significantly alter the map she created (explaining why she choose to release the Causlings when she did):

> "I already had my whole map laid out, all my islands out there, and there are a couple and I'm like 'hmm, I know these will fit in somewhere' but I already made all these connections I can't just squeeze it in. I know that it would go somewhere but I couldn't squeeze it in, in terms of the map I had already laid out. ... It's like already made...

> [thinks] Let's say you're cooking a meal and you're already done cooking the meal but you're like 'oh my gosh, I totally forgot potatoes in here' but everything is already cooked. Potatoes would go perfectly well in the dish, but you can't just throw potatoes in at the end cause they'll take forever to cook. Does that make sense? That is my problem here" (P27, University student).

Although the causal map could be adjusted while being built, players may have felt committed to particular relationships once they were established and reified (the same way that saying something out loud can influence whether you believe it (Cialdini, 2001)). Such a commitment may have encouraged a breadth-based building strategy, as players would build a portion of a map, and then once established, try to expand the map further.

Overall, players were more likely to build causal maps in a sequence resembling a depth-first traversal in terms of adding new islands and individual bridges, but build the map as a whole similar to a breadth-first traversal. Moreover, this analysis suggests that players who build maps out of linear chains may have weaker understanding of causal relationships (resulting in lower scores in the *Causlings* game), further emphasizing the importance of non-linear systems thinking in understanding causality.

6.3.3 Do players have conscious plans or strategies for constructing causal maps?

Despite the *Causlings* interactive system being framed as a game, players seemed to approach the game more as a puzzle to be solved than a game to strategize about—a valid approach given the design of the system. Indeed, the simple design of the game prototype (discussed in more detail below) did not support forming specific strategies or tactics for playing the game, in the sense of choosing the best moves to make over the course of the game (as one might think of strategy in a game such as *Chess* or *Starcraft*)[7]. The lack of clear instructions about how the

[7]In game theory, strategy refers to the choice of moves where the outcome is dependent on the moves made by another player—which of course the *Causlings* prototype does not have.

game could best be played and beaten (omitted to increase the accuracy of the interactive system as a tool for assessing of causal understanding) made it difficult for players to plan strategies in advance—though numerous players did adjust their map building style and approach across the multiple games they played. Moreover, as described in Section 6.1, players also lacked the regular feedback about how well they were doing that would enable them to properly evaluate their current strategies and form or adjust tactics while playing. In these ways, the design of the interactive system restricted the strategies that players could form or implement, regardless of the system's game framing. Nevertheless, many players were self-aware of and (when prompted) reflected on their approach to constructing causal maps and solving the puzzle of the game, even if their approach might not be considered as a strategy per se. Note that these reflections are all self-reported, and thus may suffer from reporting bias; nevertheless, they are indicative of and supported by the behaviors identified in the system's recorded interaction logs.

For instance, some players described how they approached picking islands to use and bridges to build. As mentioned previously, more than a few players reported trying to build bridges based on what was obvious or common sense—for example, one player wrote: "I tried to pick the most obvious relations, like for example, as more fish are caught, the population will go down" (P47, Turker). Others chose islands that they were familiar with, as one player explained: "I chose islands that had topics I was familiar with and that would be easy to make connections with to the samaki population (initial island)" (P7, High school student). On the other hand, some players simply used a random or brute-force approach, selecting islands as they noticed them. One player reported: "I didn't really have a particular strategy for playing the game. I added islands from the list in order and checked if the island I'm adding had any relationship with the islands already in place and built a bridge if it did" (P32, University student). Players had somewhat mixed success with these strategies (though the few players who mentioned choosing islands they were familiar with anecdotally earned slightly higher average scores), and even would change strategies over the course of playing the game, such as one who explained: "At first I just kind of did it by common sense, which didn't work because if you went back to the story, the story's really specific when it comes to the details about the fish" (P29, University student). So when presented with the initial problem of constructing a causal map, players seemed to fall back on what was intuitive—and players' intuitive understanding of causality in complex systems is often flawed.

One of the most common strategies that came up in the interviews and surveys was grouping islands into categories, as described in Section 6.2.2[8]. Many players described sorting and categorizing the islands independent of the game mode they played. For example, one player adopted a grouping strategy for his last and highest-scoring game: "The last game I tried to group them together—group the islands according to their relations. Like livestock, commercial area; and then the health, the omega 3; the larger fish; etc. I kind of grouped the islands together and made the connection that way. I just found it more efficient" (P31, University student). Categorizing the islands made constructing the causal map more efficient and easier to do, further indicating the potential benefits of pre-mapping categorization for improving causal understanding discussed above. A few players even suggested that such categorization would make the game easier and more understandable: "The islands on top were spread out in a certain way. Maybe categorizing them, for me, would have been a little easier. Like economic-based... that kind of deal" (P19, University student). A common first step in players' strategies for constructing causal maps was to categorize the islands, and so having them pre-categorized would enable and ease that strategy.

Moreover, this player's suggestion to use "economic-based" as a category reflects one of the most commonly reported categories and overall strategies for constructing the causal maps. Players in interviews explained their strategies with statements such as: "I was thinking a lot about economics and demand/supply kinds of things" (P10, University student) or "I do think if I was majoring in economics or business I thought I could do better than now" (P18, University student)—the latter indicating a player who saw an economic-focus as an optimal strategy, though not one she felt trained enough to use. When asked why she "focused on the economic things and the sale price",

[8]The frequency which people described categorized they used is possibly a side effect of the interview process. As an item of interest, I noticed such grouping and organization during my observations and so would ask the players about the phenomenon.

one player replied "I don't know; money focused I guess. It seemed like there were a lot more relationships that involved economic factors" (P28, University student). It is unclear if the prominence of the economic-focus strategy is due to the players being inherently "money-focused," or is a result of the large number of causal factors relating to economics—(Plate, 2006, Fig. 6-1) lists more than 1/3 of the system's causal factors under the economic themes of "Commercial Demand," "Samaki Industry," and "Indirect Economic Aspects." Moreover, as I've argued above, economics represents a causal system (e.g., supply and demand) that people and particularly students are trained to understand, as well as being a common framing for topics in environmental sustainability—likely providing further basis for the strategy of focusing on economic factors.

Furthermore, players' conscious strategy of grouping islands and constructing a causal map based on those groups could have encouraged the hierarchical building patterns described in Section 6.3.2. One player explained her strategy with:

> "I think if I went with a theme I tried to stick with it. Like for example if I was doing the population and the quality of the water and then the algae dead zones, I tried to stick with that instead of mixing all that with how many that the fisherman would catch. I would just stick to one theme at first and then try to do all of those related to the water quality and the weather and everything, before I would move to other ones, because I felt like if I mixed them I would get confused" (P29, University student).

This strategy of working with individual categories or "themes" describes a possible breadth-first building pattern, where each category is connected before broadening out to the next. Alternatively, such categorization could also support a depth-first building pattern, if each category is treated as an individual chain of causal relationships. Nevertheless, I did not find any players that described using a depth-first pattern to construct causal maps, while a number detailed breadth-first patterns. One player wrote: "I tried to surround the 'Capital' with variables that would directly affect the Samaki population levels, then created outer rings of variables that were more indirect effects" (P48, Turker), indicating a breadth-first "direct, then indirect" approach. Another explained "At first one caught my eye and so I would put that down... just random or whatever, maybe on the first few pages. Then I tried to find things that connect with that as I looked through the pages" (P14, University student), suggesting a "find an island that connects with what is already included" strategy that also results in breadth-first construction. As detailed above, the breadth-first strategy was correlated with higher game scores—possibly because players using this approach were more conscious of the strategy they used (so that they were more likely to reflect upon it). Players who used depth-first building patterns may have had a less consciously framed strategy, potentially explaining why this "guessing" strategy was correlated with lower game scores. Still, hierarchical building strategies were much less frequently reported than the grouping strategy, indicating that such hierarchies were not a consciously prominent component of players' game-play plans and strategies.

Indeed, executing strategies such as categorizing islands often involved careful and conscious planning. One player explained: "In the beginning I chose to work with whatever I have on the screen as fast as I could, but it turns out that it needs more knowledge and planning than that" (P14, University student), while another complained about how the time limit in the timed game mode restricted the ability to plan island groupings: "3 minutes was not enough time to read through all the islands and categorize the causal relationship between them all" (P24, University student). Other players used the grouping strategy and the organization it encouraged as part of the process for planning their maps. For example, one player described her organization with:

> "I definitely was [grouping]. I felt like they're almost like little chapters in a book if that makes sense, and then there's like the little links, so I tried making their little groups first and then making the final... I was doing the production side, and then the fish side, and then the environmental factors on the other side, and then human population, and the desire for omega-3s and fish oil that kind of was all in one section. Almost like an outline, like when you outline a paper you kind of group things together" (P37, University student).

In fact, another player recruited through MTurk reported constructing an actual outline to guide his map building: "My strategy after reading the article was to make an outline with bullet points to help me remember key points. I chose to make bridges out of the major points that I had in my outline that I could easily connect" (P46, Turker). While most players did not go to this level of detail or effort, some did acknowledge the necessity of a conscious strategy and plan of attack for the game, with the strategy of grouping islands performing almost double-duty by enabling such planning and organization.

Yet at the same time, a number of players (even the same players) reflected that their strategy involved to some extent simple trial and error. The same player who described grouping as like outlining a paper also commented on how much of the game seemed to be luck: "I feel like a lot of it is trial and error... I feel like it's all luck in a sense. I mean obviously yeah you use your intuition and like logic of causal relationships, but you just don't know how many connections should you make or anything" (P37, University student). For this player, the cognitive mapping process involved systems thinking and logically understanding the complex system, but interacting with the game-based framing of causal mapping in this interactive system—required trial and error. Indeed, players reported using trial and error for particular aspects of the game, such as one player who described how he would change the direction of bridges in reaction to the Causlings failing to cross, using "Mostly trial and error. Just to see what would work" (P30, University student). Another player applied trial and error to the strategies he used for the game, attempting to switch from building the single linear chain presented in Figure 6.5:

> "I tried to make two chains because I wanted to see if it would work cause there were all these other islands that I hadn't used at all and there were like the people's view on the fish population and the corporation, and I wanted to see if I started off a chain with those on the other way–see if that would affect it, and pretty much it was just a mortality rate on that side" (P41, University student).

The lack of clear instructions and feedback within the game framing (see below) may have contributed to some players using a trial and error approach to causal mapping, while others experimented with different strategies to try and find one that worked, without necessarily drawing heavily upon their understanding of the causal relationships in the system.

Thus overall, players seemed to be consciously aware and plan how to group islands into categories, but did not develop well-defined strategies beyond this form of organization. In this way, the game framing of the *Causlings* prototype did not seem to have a significant influence on the "tactics" used to interactive with the system and organize causal maps.

6.4 What characteristics make *Causlings* effective or ineffective as an engaging game?

Table 6.13 presents the quantitative results of questions measuring engagement from the study's closing survey. In general, the *Causlings* prototype rated moderately highly in engagement—players commonly agreed with statements that indicate the game produced a sense of flow (or were closer to agreeing than not agreeing, with all questions receiving average scores greater than 3.5 out of a 5-point Likert scale). The highest scoring variable was the game's reported ability to keep player attention, as well as their overall enjoyment of the game. These results suggest that the game was engaging in the sense that it draws players into it, and that they did not report dislike playing. However, the relatively lower agreement with the statement that the game reaches a midpoint between boredom and anxiety reflects the common perception of the game as "hard" or "challenging," suggesting that the game was not calibrated precisely enough to help players achieve a flow state. One player, in asking for a clarification on the statement while taking the survey, explained that he would get very anxious right before he released the Causlings (though he described it as a not undesirable tension) (P19, University student). In addition, agreement with the statement "I would play this game again" was also relatively low—likely due to that the players already played *Causlings* multiple time with no variation between games[9]. Building causal maps around a greater variety of content (see Chapter 9) may help improve the game's repeatability.

The lowest scoring statement across both modes asked about appropriate in-game feedback. Well-

[9]On average—though not in the median—study participants played slightly fewer consecutive games in the timed mode condition; see above and Table 6.1

Table 6.13: Engagement scores of different game modes. 5 = Strongly Agree with the statement; 1 = Strongly Disagree with the statement. There are no statistically significant differences between game modes.

	Basic Mode	Timed Mode	All Surveys
The game kept my attention.	4.32	4.57	4.41
I understood my goals throughout the game.	3.94	3.93	3.94
I received good feedback on my progress in completing the game.	3.62	3.32	3.51
I enjoyed the game without feeling bored or anxious.	3.76	3.79	3.76
I felt I could use my own strategies for completing the game.	3.97	3.71	3.86
I understand the basic ideas behind this game.	4.09	4.04	4.1
After playing the game, I feel that I better understand the topics included in the game.	4.06	3.89	4.01
I would like to know more about the topics included in this game.	3.71	3.75	3.77
I enjoyed this game.	4.18	4.18	4.14
I would play this game again.	3.74	3.96	3.87

designed video games make significant use of feedback to keep the player informed about how well they are doing, when they are on the right track, and what options are available to them. Indeed, this kind of direct and immediate feedback is one of the key components that can make games into effective learning engines (Gee, 2003a). For example, players can explore and experiment with a simulation, and receive feedback about the effects of different actions in order to learn how to play the game. However, *Causlings* was purposefully designed to offer little feedback about how well players are succeeding at the game in order to better measure their understanding of causal relationships. While in-game feedback could have helped players to build more correct bridges, it may have interfered with accurately assessing how well they understanding causality (rather than just measuring how well they can react to the Causlings). As such, players received only a small amount of game feedback: in the timed mode players could see when Causlings fell off incorrect bridges (though without indication of what was incorrect about the bridge), and in the basic mode players were only told which bridges were correct once the game was over[10]. In fact, a number

[10]Interesting, the timed mode received a lower score on feedback than the basic mode, even though the time mode

of players in basic mode would release the Causlings before completing a map, just so they could see if they were on the right track—as one player said: "I was wondering 'is this the right way to do this?' I didn't have much confidence about myself during this game" (P18, University student). In this way, this issue of designing an appropriate level of feedback is a significant concern in the development of serious games for assessment, in order to make games that are engaging and easy to play, but also remain valid assessment tools.

Although there were some slight differences between the engagement levels of the two game modes, none of these differences were statistically significant (the closest was the higher attention level of the timed mode, which had a *p*-value of 0.07). This lack of distinction suggests that the use of the Time Limit game pattern does not make a causal mapping game significantly more or less engaging. Indeed, the engagement level of a particular game mode may largely be a matter of personal preference. A number of players suggested changing the game so that it closer resembled the mode that they were *not* playing. For example, one wrote, "I think it would have made the game more interesting if the causlings were actually moving while I was trying to make associations" (P50, Mechanical Turker), suggesting gameplay like in the timed mode, while another suggested: "Ideally, if you just had a button that says 'ready' and that let the Causlings leave their first island, then you could understand 'okay, this is why it works'" (P24, University student), suggesting gameplay as in the basic mode. While the difficulty, game length, and overall timing of the timed mode requires further balancing with future development iterations (see Chapter 4), interviews suggest that at least some players believed that the interactivity of the timed mode was more engaging—and indeed this increased interactivity could help explain the slightly (but not significantly) higher level of attention kept by the timed mode.

provided more feedback in the form of Causlings falling off bridges. This difference may be due to the lack of transparency in which bridges Causlings would take (and when they would stop traveling and the game would end), or to other flaws in the design of what feedback there is. In this kind of game for assessment, poor feedback may be worse than no feedback at all.

6.4.1 Game Narrative and Scoring

Based on my observations, players seemed to engage with the narrative of the *Causlings* game prototype, as well as the interactive causal mapping system's framing as a game. I noticed numerous players smile and laugh at the opening narrative explaining the Causlings' plight, and at least three players referred to the game as "cute," suggested that the design of the story and creatures was effective. Overall players seemed to empathize with the creatures and easily anthropomorphized the abstract graphical representations. Players would react when the Causlings began to move and if any of them fell off the bridges, with one playing crying out "oh no they're dying!" when her map wasn't correct (P29, University student). One player told me: "I started to feel for these Causlings—like I don't want them to die. 33% mortality was pretty appalling to me" (P20, University student). Indeed, most players focused on the "mortality rate" component of their final score—the player appalled at his mortality rate wasn't sure he did very well, and yet achieved one of the highest game scores (based on Plate's rubric) of all the participants. And because survival rates tended to be lower than expected (often multiple Causlings would fall off of each incorrect bridge, and practically all games had at least one incorrect bridge), it seemed that many players did not believe they did particularly well at the game, even if they got a relatively high score on their causal map (e.g., high scores on the assessment). The focus on Causling survival encouraged by the game's framing and narrative diverged from my own focus on the final game score, which assessed the accuracy of the causal map. While these two metrics were correlated, it may be possible to more tightly integrate them, harnessing players' empathy for the Causlings to encourage further gameplay and interaction—or potentially to further separate them, in order to provide effective in-game feedback (i.e., Causling mortality) without invalidating the assessment metric.

In fact, one of the most prominent gamification techniques used in this system prototype—providing a score at the end—was surprisingly effective at encouraging further play and system usage (based on self-reporting). Numerous players remarked on how seeing their score at the end made them want to try again to improve it, with comments such as: "I like that it gave a score of how well I did at the end because it made me want to play it again until I could get all of the bridges correct and score 100% on the game" (P10, University student), or "Every time you see your score at the end, I think that was really what got me. Because when it first showed my success rate I was like okay lets see if I can do better, and slowly you want to incorporate more islands" (P35, University student). Other players acknowledge the accomplishment of improving one's score through multiple games: "What kept me interested was the points. You start off with negative and that makes you feel kind of bad and you want to redeem yourself, so you keep going and going" (P14, University student). Many players received negative scores during their first (and sometimes later) play-throughs, indicating that they had more incorrect bridges than correct bridges. Interestingly, a number of players were confused about receiving a negative score, asking me why there was a negative sign in front of their score; it's possible that players (most of whom were students) assumed that the lowest score you can get on something is a 0, and did not understand the idea of a score penalty for being wrong—an assumption with potential implications for the scoring design of future games for assessment.

Other players acknowledged the lack of context for the final game score (the rubric for which was purposefully left obscure), and gave suggestions for how the game's feedback might be improved. One player suggested during the survey: "There's a score, but no purpose for the score. Most games have a high score or achievement section to make the player feel good for doing well"[11] (P24, University student). She elaborated further in the interview:

> "[The score] didn't have any meaning. It was like 'okay, there's a number...' because there was no objective to the game in the beginning like 'try to score 30 points in each category and this is how its counted' or 'there's a high score at the end of this for all of the players of the game.' You just got a number with nothing to the importance of the number... it's not even like 'you win' or 'you lose'" (P24, University student).

Interestingly, this is the same kind of complaint that is attached to many descriptions of gamified

[11] Internal playtesters also pointed out the benefits of a leader board of some kind, which had been categorized as an "optional" component during development. The *Causlings* prototype lacks such a social scoring system because of time constraints on the development cycle, as well as to take extra care to preserve the privacy of participants in this IRB-approved user study.

systems (e.g., Bogost, 2011; Deterding, 2011): the points that you receive are divorced from any meaningful context, so that granting points may in fact detract from the serious purpose of the gamified system. While I did not see evidence of such crowding-out in this study, this study does highlight the importance of carefully framing how scores are presented in a game—particularly when the scores may also be tied to assessment of learning or knowledge.

6.4.2 Confusion and Instructions

While running this user study, the most common responses to *Causlings* that I received was that the game was "interesting," that it was "challenging," and that it was "confusing"—at least at first. Many players had difficulty figuring out how to play the game, despite or perhaps because of the copious instructions. The majority of players I talked to remarked on how much reading was involved—two-page pieces both to understand the causal bridges and to learn about the samaki, and then another graphic explaining the game's interface—in order to even begin playing. By the time players had finished reading about the samaki, many of them had forgotten key details about causal bridges such as the meaning of negative bridges[12]. Taken together, this abundance of text made the game's instructions unclear and confusing, though players were often able to figure out aspects of the game through repeated play. One player simply requested "Educate me better before playing" (P1, High school student), while another elaborated that: "A better explanation on causes and bridge direction would have made the initial playthroughs less confusing" (P6, High school student). Although I adapted the description of causal bridges from Plate (2006), this introduction may not have most effectively supported the samaki content drawn from the same source—one player pointed out that the example in the instructions (illustrating the effects of word-of-mouth on watching television) was "easy," but the actual game was "hard." A more abstract tutorial may have better prepared players to deal with the game's content, and enabled them to more readily engage with the causal mapping process.

[12]Due to an oversight on my part, players were not able to review the instructions on building causal bridges, though they could review the samaki article.

Indeed, many players were initially overwhelmed by the scope and quantity of causal factors (e.g., islands) used in the game prototype—36 islands split across 6 pages of the "dock" at the top of the screen. One in-person subject (P10, University student), upon reaching the main game screen and flipping through the pages of islands, purposefully turned around in her chair and gave me a look that could only be described as skeptical. Another player mentioned in the closing survey "it was lots of info to keep straight" (P44, Turker), while a third started off the interview and her description of playing the game with: "At first I was kind of scared because it was a lot of information to take in at one time, but then after a while you kind of get the hang of it" (P35, University student). Although this prototype used the list of causal factors identified by Plate, the framing of the interactive system as a game may have meant that this list was too large for engaging play. Moreover, as described above, Plate's CMAST study procedure involved the subjects picking and categorizing causal factors before they began to construct causal maps; as this step was not included in my study (though many players performed the such categorization voluntarily; see above), players didn't have a chance to familiarize themselves with the content before beginning play—an absence particularly felt in the timed mode. It is unclear whether this abundance of choices would continue to be a problem when assessing causal understanding that is not based on a prior reading, as with a non-prototype form of the game. Nevertheless, players were often able to figure out the game after a try or two and seeing how things played out—thus it is possible that more concrete examples or tutorials may lead to less confusion and improve engagement (though see also Andersen et al., 2012).

Despite the steep learning curve and initial confusion, players overall report enjoying and being challenged by the game—as one player summed up in the closing survey: "it was a challenge at first I didn't understand it at all and I grew to like the game" (P52, Turker). Indeed, some players were very excited about the game—one referred to it as "the future of learning" (P19, University student), while another (P20, University student) talked about letting his 6-year-old nephew play the game (though feared the competition for the best score that it might entail). Overall, the gamified elements of even this prototype system were moderately effective at engaging players, and encouraging them to participate in assessing their causal understanding, though the game framing also placed limits on the interface's complexity so that the game could remain engaging.

6.5 Conclusion

In this chapter, I analyzed the results of the user study evaluating the *Causlings* prototype, considering how the design of the system and its framing as a game influences its use a cognitive-mapping based assessment of causal understanding, as well as offering an initial exploration of the process by which players construct causal maps.

I found that while the particular interaction modes (timed and basic) did not have a significant effect on the system's assessment, the game presentation and interface did have a significant impact on the causal maps that players constructed—in particular the interface likely led players to primarily include in their maps islands that were available on the first page of the dock. I also found that as an assessment, the *Causlings* prototype measured players' causal understanding to be potentially less accurate than that found by previous research (being measured with a slightly different scoring rubric), though the maps constructed had equal levels of "web-like-ness" and interconnectivity. I suggest that this discrepancy is due to the lack of an explicit categorization step in the cognitive mapping process, as evidenced by the desire of players to perform such categorization and the anecdotally higher scores of those who did.

In terms of what the prototype revealed about the process by which players reify causal maps, I found that players frequently organized the islands into categories as part of a conscious strategy for organizing causal maps. In particular, players would focus on the group of islands related to economic impact, drawing on an economic interpretation of causality and the sustainability of the causal system. But beyond this regular grouping, there were few visible patterns in the sequence of islands used or bridges built. This result potentially suggests that the sequence by which a player constructs a causal map is highly individual, and thus studies of this sequence may shed further light on whether and how an individual understands causal relationships. Nevertheless, players were slightly more likely to build following a "depth-first" approach (constructing linear chains of causality rather than branching stars), though this approach was negatively correlated with the game score—as documented in the literature, players who relied on linear chains had weaker measured understanding of the system's causality. At the same time, a "breadth-first" approach was slightly linked to high game scores, potentially indicating the benefits of an iterative, constructionist approach (e.g., Papert, 1980) to learning about causality—the more players can construct categorized mental models and build these models off of one another, the better.

As a game, the *Causlings* prototype was moderately engaging (particularly the game's narrative), though the instructions were often confusing and the variety of choices overwhelming. This result suggests that the overall frame may be effective at encouraging assessment of causal understanding, though the implementation could be improved to better measure understanding of a wider variety of causal systems.

In sum, the interactive system is able to achieve its objective of enabling the automated assessment of causal understanding through cognitive causal mapping, and is able to do so in the form of a moderately engaging experience. Moreover, this automated system enabled me to begin the exploration of the cognitive mapping process and the sequence by which players construct causal maps, using this sequence to further evaluate their causal understanding. In the next chapter, I further discuss the implications, contributions, and significance of these results and of my research as a whole.

Chapter 7. Discussion

In this dissertation, I consider how an interactive computer system in the form of a game can support assessing people's understanding of the causal relationships in complex systems through the process of cognitive causal mapping. The results of the previous chapter suggests that this system is able to automatically apply such a causal mapping assessment technique through the computer medium. The results also demonstrate that the system can enable assessors and the system itself to better consider the resulting form and sequence by which players construct causal maps that reify their causal knowledge. However, the system's interface and game framing may have introduced some side effects that influenced player's causal mapping, raising questions about the influence and appropriateness of gamification in an assessment context.

In this chapter, I discuss some of the broader implications of these results. I further analyze the results of the assessment, considering what the form of causal maps and the sequence by which they were built may reveal about causal understanding and the systems thinking process as a whole. I then discuss some of the effects of the system's game framing on the assessment process, identifying characteristics of this and similar games that may make them less suitable for supporting stand-alone assessment. I next detail the key contributions of this dissertation, justifying these contributions based on the presented research. Finally, I discuss some of the broader impacts of these contributions, particularly their applicability to both formal and non-formal educational contexts.

7.1 Causal Map Construction Reveals Failure of Systems Thinking

In considering the maps constructed through the *Causlings* system during the evaluative user study, the structural forms taken by these maps—particularly players' use of categorization—suggests a highly reductionist approach to understanding the causal system, while the sequence of map construction shows an emphasis on linear thinking. Both of these approaches are frequently positioned as opposing the kinds of system thinking required for understanding complex causal systems; thus the forms and sequences of causal mapping created by many users of the *Causlings* system suggest a failure of systems thinking.

7.1.1 Structural Forms Suggest Reductionist Thinking

In the evaluation of the *Causlings* system, many players self-reported being overwhelmed by the scope and complexity of the causal system, as introduced by the large number of factors in the system and the difficulty of considering the interactions between them. In order to deal with this overwhelming complexity, *Causlings* players would attempt to break the system down into manageable components, primarily through some form of categorization. As described in the previous chapter, many players grouped islands into different themes or categories, and still others requested that the islands be initially categorized in order to make the game less confusing (that is, easier to understand). Furthermore, in comparison with Plate's (2006) study that did include an explicit categorization procedure, *Causlings* players received lower scores (even after accounting for discrepancies in scoring rubrics), suggesting that this categorization improved the assessment of their causal understanding. In these ways, the dissertation results suggest that categorization is an important component (at least in players' perception) of making sense of complex systems and constructing accurate causal maps. Anecdotally, players who categorized the factors of the causal system received higher game scores, indicating a higher level of understanding.

Why was categorizing causal factors such a common and effective strategy among players? I interpret this prevalence as resulting from players having trouble holistically understanding the complex system, and so attempting to simplify and reduce the available choices in order to make decisions (e.g., about which bridges to build). Players appear to desire to reduce an overwhelming complex system to one whose complexity they can understand—to shrink the scope of the system to a size that can be cognitively processed. Indeed, the human brain has not evolved to easily comprehend large-scale complex systems, such as those dealing with issues in environmental sustainability (Tomlinson, 2010). Players could not holistically understand the complex system surrounding the samaki, and so tried to break the system down into its components through grouping and categorization, which players could then deal with logically (using "common sense" to build bridges, as many reported). Indeed, the frequency of islands used can be understood as players compartmentalizing the factors using simple partitioning—considering the subsystem made up of just the first islands listed.

In this interpretation, the results of this dissertation suggest that players deal with the overwhelming complexity by restricting the size of the system that is currently under consideration—they reduce the problem of constructing a causal map to deal with a smaller, less complex system that they are able to understand intuitively. As such, categorization can be seen as an example of *reductionist thinking* (see Section 2.2.2). Players reduce the system into component parts (themes or categories) that can be addressed and reified individually, and then attempt to combine these subsystems into a holistic understanding (often with only a single relationship connecting them, as in Figure 6.6). This dissertation suggests that such categorization—and the reductionist thinking it represents—can be identified through a consideration of the structure and the spatial layout of the reified causal maps. Indeed, further extensions to this prototype that apply existing techniques for automatically detecting the degree of graph clustering (e.g., Schaeffer, 2007) could potentially enable an interactive system such as *Causlings* to automatically detect levels of reductionist thinking, which could further inform an assessment of systems thinking skills and causal understanding. For example, a player who reifies causal maps with a high degree of categorization or clustering during this assessment may benefit from further education in holistic, systems thinking approaches to complement his or her current use of reductionism.

It is worth noting that the CMAST causal mapping technique (Plate, 2006) upon which the *Causlings* assessment is based is itself a somewhat inherently reductionist approach to reifying causal understanding. In this method, subjects are asked to consider each combination of factors pairwise, in isolation from other relationships. Subjects reduce their consideration to a single pair of factors at a time, rather than using a more holistic approach to thinking about the system. While the physical construction of the reified causal map (i.e., laying out the factors on paper and drawing the lines between them) can potentially invoke more holistic, systems thinking considerations, the primary interaction with the CMAST technique reduces the system to a collection of independent causal relationships that can then be combined—the very essence of reductionist thinking. Thus simply using CMAST or a similar method as a technique for assessing causal understanding may encourage reductionist approaches at the expense of systems thinking. But at the same time, it is possible that as a graphical interactive system, *Causlings*'s emphasis on the spatial layout of reified maps might encourage more holistic approaches to causal mapping. Players may be encouraged to consider how each new factor fits into the map as a whole (i.e., where to fit it in to the spatial layout), rather than only thinking about a single pairwise relationship. Studying whether such interactive systems encourage reductionist or systems thinking approaches to constructing causal maps is important future work.

In applying reductionist thinking and breaking the causal system into its component parts, one of the most common component parts of the causal system that players considered was the part of the system involving economic factors. The results presented in the previous chapter indicate how often players consciously included and explicitly grouped economic causal factors—indeed, many of the most frequently built (and most frequently correct) bridges represented economic-based causal relationships. As I mentioned previously, discussions of environmental sustainability are often framed in terms of markets and economics—carbon footprint calculators are tightly linked to purchasing offsets (Ross et al., 2010c), and pushes for energy use reduction or vehicle fuel efficiency are pitched as saving people money on electricity and gas bills (see also Dillahunt et al., 2009). (Dourish, 2010) notes how sustainability is often framed as "an instance of an economic rationality of costs and benefits," with people readily applying market models to consideration of environmental impact. The high frequency with which *Causlings* players included economic-based factors in their constructed maps and conscious categorizations—indeed, the prominence of economic concepts in the factors chosen by experts to be included in the cognitive mapping exercise—highlights the pervasiveness of this market-based view in considerations of sustainability (and potentially of complex systems in general).

Moreover, Dourish also explains how the economic consideration of sustainability focuses on *individual* choice and consumption, following the market logic that individual decisions and behavior change will aggregate into collective environmental reform. Considering Dourish's insight in light of the above discussion, one can see the focus on individual choice as itself a similar form of reductionism: people try to influence the complex system and support environmental sustainability by influencing a single component part (namely their own behavior or consumption). Yet as I and others have argued, a reductionist approach is not an effective method for understanding complex systems such as sustainability—with this approach people may miss the more influential and even necessary strategies for environmental sustainability (such as large-scale political or cultural change) that might be more readily apparent from a holistic, systems view. People often don't think about trying to change technical or social infrastructures, because in a reductionist view (the prevalence of which is supported by this dissertation) they fail to see the significant emergent effects of those factors. In this way, by considering the economic relationships between causal factors, many players missed (or incorrectly understood) the social and cultural factors that are part of the game's complex causal system. The results of this dissertation thus reiterate the prominence of economics in how people understand the complex system of environmental sustainability—a reductionist approach that leads people to have difficulties understanding and effectively interacting with this massive-scale complex system. Further analysis of the common forms of reductionist thinking—such as what are common groups of factors, as can be automatically detected by an interactive system such as *Causlings*—may be able to inform further educational efforts in how people can effectively interact with complex systems such as environmental sustainability.

7.1.2 Construction Sequence Suggests Linear Thinking

The results of this dissertation's evaluative study also demonstrate an algorithmic technique used to analyze the sequence of causal map construction—in particular, whether maps were built in a more breadth-first or depth-first manner. Furthermore, I suggest that a depth-first approach may reflect an emphasis on building chains of causal relationships (where each new factor is connected to the previous in the chain), while a breadth-first search may reflect an emphasis more on the map as a whole (as previously-placed factors are connected broadly to a wide range of new factors). Because a causal chain consists of a linear sequence of causal relationships, I interpret depth-first, chain-based map construction as an indication of a form of *linear thinking* (see Section 2.2.2). Players may be considering the relationships in the causal system as a linear series of connected steps, focusing on how the first relationship leads to the second leads to the third. This is the essence of linear thinking; causal chains are the expected organizational result of a linear thinking approach to reifying a causal system. Thus players who demonstrate high degrees of depth-first map construction—constructing their maps through the organization of causal chains—may be demonstrating a prominently linear thinking approach (as opposed to a systems thinking approach). Along the same lines, I interpret a breadth-first sequence of causal map construction as representing a more systems thinking-based approach, with players interacting with the previously reified map holistically rather than reducing it to linear chains. In this view, depth-first sequence building indicates linear thinking, as opposed to breadth-first sequence building that is more likely to indicate systems thinking.

Based on this interpretation, the interactive system prototype presented in this dissertation is able to automatically detect linear thinking in the construction of causal maps using the analysis methods detailed in Chapter 6. The sequence by which causal understanding is reified can be computationally parsed to determine the degree of depth-first building, which can indicate a linear thinking approach to understanding the complex system. In this way, the presented system offers a novel tool for assessing a player's use of linear thinking as opposed to systems thinking.

Moreover, the results of this dissertation suggest that a depth-first, linear thinking approach to reifying causal maps was not effective in understanding complex systems. Depth-first building was significantly and negatively correlated with game score; I interpret this result to mean that players who were more likely to apply linear thinking approaches were less able to identify causal relationships in a complex system, suggesting a less complete or accurate understanding of the system. Linear thinking was less able to help (and may even have interfered) with the systems thinking required to understand and interact with the complex system—and this linear thinking could be detected through a consideration of the sequence of causal understanding reification. This evaluation also suggests that such linear thinking was a highly common approach to trying to understand and reify the causal maps—it just was not an effective strategy. At the same time, following one of the three metrics described in Section 6.3.2, breadth-first building was positively correlated with game score, suggesting that the more holistic, systems thinking approach it represents may have been more effective in understanding the causal system. However, this result is less conclusive; while the system is able to detect linear thinking, further study on how it may detect systems thinking is required.

Note that a reliance on the construction of linear chains in a depth-first sequence can also be seen a form of reductionist thinking, as well as a type of linear thinking. Creating individual causal chains can be interpreted as players focusing on a single topic of sequence of causation at a time, rather than a more holistic consideration of the interconnected relationships in the causal system. In this way, a linear-thinking approach is related to a reductionist approach, both of which are in contrast to the systems thinking skills required to understand complex systems. As a form of reductionism, chain-based linear thinking may be less effective at understanding the many emergent properties of complex systems for the reasons described above—thus linear thinking may be negatively correlated with game score simply because it also represents a form of reductionist thinking.

The high frequency of a depth-first, linear thinking approach to constructing causal maps is not surprising, giving the prominence of linear narratives in human cognition and traditional science education (see Section 2.2.2). Linear, depth-first building may represent players attempting to construct a story of how factors in the causal system are related. Indeed, some players (who misunderstood the structure of the system's game framing) created linear causal chains that explicitly represented the samaki's life story–from birth to capture by humans to becoming a source of omega-3. Moreover, this inclination towards a linear, narrative-based approach to understanding the causal system may have been exacerbate by the narrative presentation of the article and of the game framing itself. By framing the *Causlings* system with a game narrative (one in which the player helps the Causlings to move to new islands), the system's presentation may have encouraged players to consider the system in terms of narrative. However, the game's narrative was intentionally developed to be one of broad exploration ("reach as many islands as possible") rather than linear travel (e.g., "go to this particular island"), with the hope that this narrative would encourage players to *deviate* from the more commonly used linear narratives.

Indeed, further work could involve developing teaching tools in the form of variations on the *Causlings* system that emphasize non-linear sequences of causal map reification. For example, the game could be modified so that bridges could only be created that connected islands not recently used, removing the ability to directly construct a linear causal chain and hopefully encouraging a more holistic consideration of the causal system and its map representation. Such a variant would try and force players to perform a kind of "multi-task" in their consideration, thinking about multiple vectors of causation simultaneously.

In fact, such multi-tasking may be a way of addressing linear thinking. Modifying existing, entrenched habits of linear thinking may require breaking away from interactions in which the user only has a single point of contact at a time. In many computational systems (including most computer games), users can only apply a single input at a time: typing one key at a time, clicking one mouse at a time, hitting one button at a time. In effect, such systems reduce the user to a linear sequence of interaction—thus understanding how to interact with these systems lends itself to a linear thinking approach in which the user considers the step-by-step sequence of inputs. Simply using most computer systems with a serial input format reinforces a linear thinking approach. This interaction paradigm was established in response to early limitations of computer capabilities and infrastructure; however, a variety of research efforts in ubiquitous computing and the development of novel modes of interaction are working to overcome these limitations and potentially enable parallel interactions instead of just a linear sequence of inputs. As such, these emerging interaction techniques may be required for supporting the education and assessment of systems thinking, without implicitly encouraging linear thinking by restricting interaction with the assessment to a sequential process.

Despite the linear nature of the system's interaction, the *Causlings* prototype is able to automatically detect the presence of linear thinking by analyzing the sequence of causal map construction for depth-first building patterns. Nevertheless, the prominence of such linear thinking may be significantly influenced by the linear narratives and interactions of the computer system and it's game framing; I discuss the implications of this framing further in the next section.

7.2 Games May Be Less Effective For Stand-Alone Assessment

As discussed in Chapter 3, a number of scholars have positioned computer games as a form of interactive system that can be highly effective at supporting learning and education. Good games (à la Gee, 2003a) represent highly *learnable* systems: well-designed games are structured to provide steadily increasing challenges to players, keeping them engaged in the process and continuously developing new skills and understanding. This development is supported by providing "just-in-time" information and constant feedback to players about their performance and how successful they are in the game. Such feedback supports the way in which games provide a kind of separate space or context (see Salen and Zimmerman, 2003) in which players are free to experiment and "play" with the game system, thereby learning and internalizing its rules through active interaction and in a sequence best suited to students' individual needs. Moreover, games can provide significant levels of intrinsic motivation—whether in the flow-like engagement afforded by dynamic difficulty, in the fun of learning and gaining mastery (Koster, 2005), or potentially in the social validation that comes from playing large-scale multi-player games (such as may partially drive the success of games like *Foldit*). Overall, games' learnability, feedback, and free experimentation makes them potentially very effective tools for education.

However, in exploring how *Causlings* as an interactive system might used for assessment (which is a significant component of education), I found that the system's game framing was frequently at odds with the system's use as a stand-alone assessment. *Causlings* was not especially effective at enabling the forms of education support afforded by games. Indeed, the very requirement of being a stand-alone assessment excluded some of the above benefits, suggesting that game framings such as that used by *Causlings* may be less suitable in interactive systems for assessing existing knowledge than in interactive systems for teaching new knowledge.

7.2.1 Playing a Game Requires Understanding the Complex Game System

Gee (2003a) refers to games as "learning engines," in that they train the player over time to master the complex problem(s) presented by the game. In order to successfully play and win at a game, a player needs to learn to understand the system underlying that game (Salen and Zimmerman, 2003)—without understanding the rules of play or what different game moves represent, the player is unable to effectively interact with the game. Video games' strength as learning engines is in how effectively they enable players to master them and learn the underlying system. Yet supporting this easy mastery requires non-trivial design efforts in order to make the game or system learnable (Grossman et al., 2009)—either through in-game tutorials or on-demand help (see Andersen et al., 2012).

The results of this dissertation's user study show that for many players, *Causlings* failed the design objective of being learnable. As described in Section 6.4.2, a number of players misunderstood the objective of the game, and even more remarked on the instructions' lack of clarity. The non-interactive, text-based instructions were not sufficient to teach many players the system underlying the game-based assessment. Players' difficulty in understanding the intricacies of the interactive game system arises in part from the complexity of the cognitive causal mapping process itself. Creating a cognitive causal map can be confusing, requiring the subject to understand clearly the structure and form of this (not necessarily intuitive) visual representation. In fact, Plate excluded influence direction (whether a bridge is increasing or decreasing) from his analysis because he didn't feel that participants understood this aspect of the causal map representation (Plate, personal communication, March 12, 2012)—the concept of influence direction (even with its analogs to correlation) may require some practice to effectively comprehend. And the game framing of the *Causlings* prototype makes the process of cognitive mapping into an even *more* complex system, introducing a number of components that can produce emergent meaning depending on how they are combined: Causlings, islands, "causal bridges", "controlling concepts", and a number of mouse-based input commands. Players needed to understand the symbolic meanings of the game icons (islands, bridges, etc), as well as the Causlings' emergent interactions with the constructed maps, in order to succeed at the game. Overall, this diversity of elements and interactions may have made the game of constructing a causal map into a confusing complex system—one which players had to comprehend in order to play and have their causal understanding assessed.

In this way, the *Causlings* prototype's game framing makes it into a complex system that players need to master to show their understanding of *another* complex causal system (i.e., the samaki) being considered. Demonstrating mastery of the complex system being assessed requires first demonstrating mastery of the additional complex system that is the assessment tool. Thus the prototype consider in this dissertation suggests that in order to use games for stand-alone assessment, players need to have an *a priori* understanding of the complex system that is the assessment. Such understanding requires further time and effort from both the subject and the assessor that is likely not appropriate or desirable for a stand-alone assessment, as well as introducing an additional point of failure for the assessment process. As such, the way that gamified assessments produce complex systems that need to be understood themselves may make games less suitable for assessment than for other aspects of education. While the narrative game framing of the Causlings and their plight was commonly able to produce buy-in and engagement (one of the main benefits cited in the application of serious games and gamification), needing one complex system to assess understanding another may make such framing too costly in terms of player (and developer) effort to be effective at supporting stand-alone assessments of causal understanding.

On the other hand, not all gamification techniques may need to introduce complex systems that require distinct mastery in order for assessment to occur. Simpler forms of gamification such as those referred to as "pointification"—the points and badges and other elements that gamification's detractors object to—may allow game dynamics to be added to a system without creating a fully new complex system that the player needs to learn and master[1]. Calling a test score "points" is easy to understand; understanding how to play a test-as-game is much more difficult. And such simple dynamics may still be able to support the kind of engagement that is one of the largest benefits of using game dynamics in serious games and games for education. Indeed, as described in Chapter 6, the points given at the end of the game may have introduced just as much incentive to continue playing as other game components. Yet in developing *Causlings*, I believed that just as measuring learning may need to be more than simply answering questions on a test, assessing knowledge through a game may need to be more than just giving points, requiring the complexity of a fully designed serious game. Nevertheless the results of this study suggest that requiring

[1] In fact, this lack of complexity is one reason that many do not consider gamified systems to be true games. If it doesn't require learning or mastery, can it still be considered a game?

players to learn one complex system in order to assess their mastery of another may be problematic. Including game elements without framing the assessment as a full-fledged complex game that needs to be mastered may be more effective and lead to less confusion (though with potentially diminished engagement) in developing stand alone assessment tools. On the other hand, as the cognitive mapping assessment procedure itself has a notable learning curve, learning to play a game such as *Causlings* may not be any more difficult than learning to perform the assessment procedure in the first place. The exact impact of particular game mechanics and learning strategies on learnability—and how such learnability may fit into an educational assessment context—required further research.

Overall, the difficulty in understanding the game framing of the *Causlings* interactive prototype points at how the complex systems that underlie games may make this form of gamification less effective in supporting stand-alone assessments. However, simpler game dynamics such as pointification may be able to support engagement without increasing the amount of the additional mastery required to perform the assessment in the first place. But even if the game framing of an assessment was intuitive enough (or relevant enough) to not require additional training, the educational contexts in which assessments usually appear offers their own restrictions on the effectiveness of game-based assessment.

7.2.2 Assessment Contexts Do Not Condone Playful Feedback

Labeling an activity as a game introduces a number of interaction assumptions for potential participants. Being a game implies that the activity exists within a "magic circle" (Huizinga, 1955; Salen and Zimmerman, 2003) in which particular rules and assumptions apply—where actions both gain and lose semantic meanings in comparison to the context outside of the game. For example, this separation (though often blurry) commonly suggests that players are free to "play" with the system: to perform moves and actions that do not have a cost of failure outside the context of the game, and thus players may interact with the game system in ways that they wouldn't with a non-game system. Moreover, for many people interacting with an activity called a game implies that the activity will provide particular forms of motivation—namely intrinsic motivations such as "fun" rather than extrinsic motivations such as money (Ryan and Deci, 2000). Indeed, the dominance of particular sources of motivation can shift an activity from being "play" to being "work" and vice versa. The purpose (e.g., to entertain) behind an activity or game can also have significant impact on how players approach and interact with the system. While a game framing suggests certain meanings and values for the interaction, these implied values may be less appropriate (or even in opposition to) the values inherent in traditional assessment contexts.

For example, consider the role of regular *feedback* in games. As Gee (2003a) explains, good games offer constant feedback, supporting players as they interact with and learn the underlying complex system that makes up the rules to the game. Indeed, this feedback helps keep players engaged with and enjoying the game as they see immediate effects of their actions and so can dramatically execute their developing mastery over the system. Yet traditional assessment contexts (e.g., tests or exams) do not support this kind of constant feedback—indeed, most forms of examination-based assessment would lose their validity if students were able to check if the answers they wrote down were correct, changing their answers before moving on[2] For this reason, in order to support the goal of assessment, the *Causlings* game does not provide sufficient (or indeed, almost any) feedback on player moves. Thus according to Gee's criteria, the interactive system is *not* a "well-designed game"—a status that was in part borne out by the confusion regarding and only moderate engagement with the game. The results of this dissertation thus suggest that a gaming context (which requires constant feedback) may not be readily combined with an assessment context (which traditionally intentionally precludes immediate feedback). These two contexts have conflicting requirements in terms of feedback that may make them incompatible.

Note that the design of the study presented in this dissertation encouraged, even required play-

[2]Considering the impacts such feedback may have on test validity reminds us of the importance of training students to be able to check their own work, in a way providing their own feedback.

ers to play repeated games of *Causlings*. This repeated play may have enabled a limited form of feedback—players would construct a causal map, see how accurate it was (as much as possible given that the short animation that did not provide significant time to analyze a map's accuracy), and then potentially use that feedback in the construction of their next causal map. The analog to this form of interaction in a traditional assessment context would be to have students take an exam over and over again until they get all the answers right. I expect many teachers would initially balk at this premise, as it runs counter to many historical assessment formats. Yet the appropriateness of such an interaction is dependent on the *purpose* of the exam. In education, assessment can have one or both of two main goals: accountability (e.g., whether students actually learned anything) and improvement (e.g., using assessment to help students learn more)—with many educators, including myself, favoring the goal of improving learning (see e.g., Angelo, 1999). In this view, the learning is the main goal—so if taking the test over again helps the student better learn the concept, then repeated attempts is acceptable (though of course, assessments still need to be fair and valid between students, and fulfill the socially- and culturally-dictated requirements of the educational system in which they are used). Thus it may still be possible to use games for assessment if the overall goal of the assessment is reconsidered—switching from only measuring knowledge to also developing knowledge. However, for stand-alone assessments whose purpose is accountability (like the interactive system developed for this dissertation), games may not be a suitable form of interaction.

Games' property of providing constant feedback is tightly related to their defining structure of providing "unnecessary obstacles" (Suits, 1978) for players to overcome: feedback supports hows games give players the freedom to experiment and "play" with the complex game system. In many video games, players can hit random buttons or try interacting with virtual objects in different ways in order to learn how the system works and how the game can be played most effectively; the constant feedback offered by games lets players know the results of particular interactions. Moreover, players are allowed to fail interactions without fear of significant consequences—often the game will simply tell the player they were wrong and the player can try again. This idea of being free to experiment in a game—of games existing within the magic circle—is a key component of games as an interactive form. Indeed, the idea of players experimenting and discovering their own understanding is one of the main principals of good learning Gee (2003a) finds in video games, while Deterding (2011) calls out this form of autonomy ("the freedom of a sandbox to have a space to play in and something that can be played with") as one of the key ingredients for games that gamification efforts often miss. In a game, players can experiment with actions and receive immediate feedback on whether they were successful, without the status of that success having significant impact outside of the game's context. Indeed, the results described in Chapter 6 suggest that many players did experiment with the system and building different structures of causal maps: whether they methodically tested new bridges to produce a high scoring map in the end, or whether such experimentation led their final game to have a lower score than previous attempts.

However, these experimentation may have been possible because as a prototype of an interactive system, *Causlings* did not have the additional connotations introduced by assessment in a formal educational context. In such contexts, assessment results are perceived to have high (potentially *too* high) value outside of the educational system. Exams are tightly linked to grades, and for many students and educators grades are the all-important result of education. Thus in traditional educational contexts, assessments are not situations in which a student feels empowered to experiment and "play." Failing an assessment usually has consequences, unlike failing at a game (where you can try again with little to no cost). In this way, the playful experimentation in which results have little meaning outside of the game represents a significantly different value system than that associated with assessment. Indeed, grades can be seen as a form of extrinsic motivation for learning—and as mentioned above, extrinsic motivations can often crowd-out the intrinsic "fun" of a system that makes it into a game. Again, this contrast arises from the purpose of assessment: if an assessment's goal is purely extrinsic (e.g., accountability), then it may not be compatible with the intrinsic motivations necessary for game-like interactions. In this way, unlike games and education, the value systems that underlie games and traditional assessment may be fundamentally opposed, causing game systems to be less appropriate for assessment contexts. Successfully framing an assessment as a game may not be possible within existing educational systems.

Furthermore, assessment games may be less suitable for ready adoption on a large scale than other forms of serious games or games for education because of the particular purpose of assessment. As discussed in Chapter 3, the *Causlings* game prototype is inspired by exemplary serious games such as the *ESP Game* and *Foldit*. These games were able to successfully reframe work as "play," and were subsequently played by hundreds of thousands of players. Yet the adoption of these games (particularly *Foldit*) may be due in part to factors and motivations external to the game: by playing *Foldit*, players can feel like they are part of a larger movement to improve the world by helping solve the problem of protein folding. In this way, players may be motivated to play *Foldit* by a kind of altruism and social validation—both forms of intrinsic motivation. Although as a serious game, the game has a purpose other than simply being entertaining, this purpose does not directly benefit the player, and thus does not act as an extrinsic motivation that may crowd-out the fun of the game. The non-entertainment purposes of the serious game are kept external to the player. However, in stand-alone assessment systems such as *Causlings*, the non-entertainment goals (measurement of knowledge) are in many ways highly *internal* to the player: after all, the system's purpose is to inform the player or the assessor how well the player understands a concept. This contrast represents yet another way in which the purpose of a stand-alone assessment may conflict with the motivations for playing games, making the addition of game framings a less effective technique for supporting assessment.

Overall, this dissertation suggests that although games may be able to increase engagement with an activity designed to assess knowledge such as causal understanding, traditional assessment contexts may not be appropriate for such game-based interventions. The assumptions and values associated with assessment contexts may prevent game systems from effectively establishing a magic circle in which the player can receive feedback on their actions in freely experimenting with the game system. While more formative or learning-driven assessments (as opposed to accounting-driven assessments) may be able to benefit from game framings and dynamics[3], the results of this dissertation suggest that the kinds of stand-alone assessments considered may not be an appropriate domain for gamification.

7.3 Contributions

In this dissertation, I explored how interactive computer systems (particularly computer games) can support assessing understanding of complex systems through consideration of the process of causal mapping. In this section I present the main contributions from this exploration:

1. *Procedurally evaluating the structural forms of causal maps created in an interactive system might suggest the presence of reductionist thinking.*

In this work, I considered the graph structure and spatial layout of causal maps that subjects created as reifications of their understanding of a causal system. Informed by subjects' self-reports, I found that these graphs often involved a high level of explicit categorization. I argue that such categorization represents a form of reductionist thinking, in that subjects reduce the map to more manageable categories that can then be recombined. This interpretation suggests that algorithmic analysis of causal maps (e.g., using cluster analysis) may be able to detect the presence and degree of reductionist thinking simply through a consideration of the structure of the map. Such analysis may hypothetically be automated as part of a computerized assessment system, though a demonstration of this process is not presented in this work. Detecting reductionist approaches in this way is an important first step in identifying a lack of systems thinking that can be reinforced through education, as well as for potentially determining the imaginary boundaries that people may cognitively place upon a system and thereby mask an awareness of emergent properties. This dissertation thus offers a potential link between map structure and a cognitive process (reductionist

[3]Indeed, the ability to easily repeat the assessment (replay the game) with potentially different content or options is one of the key benefits of implementing the CMAST causal mapping assessment technique as an interactive computational system.

thinking) that may be measured through the presented interactive system for assessment.

2. *The sequence in which a user constructs a causal map can be analyzed to detect the presence of linear thinking, and this analysis can be automated by an interactive computer system.*

In this dissertation I offer a novel exploration of the *sequence* by which subjects construct causal maps through assessment techniques such as CMAST (Plate, 2006). With this work I present an interactive system that is able to automatically record this sequence, enabling further analysis. I then demonstrate a programmatic analysis of the construction sequence, one that can be performed in real-time when integrated with a computer-based assessment tool. This analysis considers whether the map was built in a breadth-first or depth-first manner—the results of this dissertation show that while depth-first building was more common, it was also significantly and negatively correlated with map accuracy, suggesting that a depth-first approach represents a less effective approach to understanding causal maps. I argue that such a depth-first sequence is indicative of the subject's linear thinking. Such linear thinking emphasizes the existence of causal chains, rather than the causal web that must be considered to identify emergent properties of complex systems. Thus the tool presented here is able to automatically detect linear thinking in the sequence of causal map construction—a process that can further inform assessments of people's system thinking skills and understanding of complex causal systems.

These two contributions—in conjunction with interactive systems similar to the prototype discussed here—may be used by educators to support their assessments of systems thinking pedagogies or of students' understanding of particular causal systems. These contributions enable causal maps to be analyzed for novel implications using a mix of old and new analysis techniques. Furthermore, the automated nature of this analysis process can also enable the assessment of large populations (as suggested in Chapter 8).

3. *Although games as interactive computer systems can be effective in enabling learning, they may be less readily effective in supporting stand-alone assessments due to (a) requiring an* a

priori *understanding of the complex game system used in the assessment, and (b) traditional educational assessment contexts not supporting the forms of feedback critical to supporting game-based learning.*

Finally, in this dissertation I offer an in-depth discussion of the suitability of game dynamics for use in stand-alone assessments (such as supported by the presented interactive game system). I find that while previous research has established the benefits of using games for education, these benefits do not readily translate to the portion of education that is assessment.

In testing the presented game-based assessment, I discovered that the complexity of the game system may interfere with the overall goal of measuring understanding. While game systems can encourage learning by persuading players to develop an understanding of the system, this learning step may not be appropriate for assessments where the goal is accountability rather than further learning. Using the presented study as an example, I argue that requiring a player to understanding a complex game system in order to participate in the assessment can interfere with measures of understanding, as (for example) inaccuracies in a causal map may be either the result of not understanding the causal system in question, or may be the result of not understanding the game used to assess that understanding. This dissertation thus suggests that the difficulties of using a complex system to assess understanding of a different complex system do not outweigh any engagement benefits of gamification.

Furthermore, in considering the relationship between games and assessment, I argue that certain elements that make games engaging and effective for education in fact make them incompatible with traditional assessment contexts. In particular, good games require providing constant player feedback, so that players are able to effectively explore and learn to master the game engine. However, traditional assessments that focuses on accountability preclude subject feedback, as this would invalidate the assessment. The external valuations placed on traditional assessment contexts (e.g., the importance of a grade) can also interfere with the construction of a "magic circle," which is needed for an activity to be considered a game. I argue that games and assessment are opposed along a number of critical axes, making games less suitable for use in stand-alone assessments.

Overall, this dissertation thus argues that while games may be useful for education, they may be less effective in supporting stand-alone assessments that focus on accountability. However, games may still be effective in supporting assessments that are combined with other learning systems (such as in projects like those described in Chapter 3; e.g., Sliney and Murphy, 2011); the nature of this relationship remains a subject of continued future work.

7.4 Broader Impacts

The system and analysis techniques presented in this dissertation and encapsulated in the above conclusions can be adopted by educators and other researchers to add further nuances to the analysis of causal understandings of systems thinking skills. Furthermore, in this section I consider in more detail a particular aspect of this research that I feel is particularly relevant to educators: the role of narrative in systems thinking education.

One of the most significant component of the game framing used in the *Causlings* interactive system is the game narrative. Indeed, the narrative is what makes *Causlings* a game, rather than simply a computerized version of the CMAST technique. This narrative converts causal factors to islands, causal relationships to bridges, and the reification of a causal map into an attempt to help imaginary virtual creatures explore. Without the narrative, the system's other "game-like" elements (e.g., the high score, the Time Limit pattern) lose their game-like qualities, and in fact would fit right in with a traditional exam (which also provides a "score" at the end, and which usually have a time minute—though often more than a few minutes!) In this way, the narrative is the fundamental game element of the *Causlings* interactive system.

Nevertheless, narratives are not required for a system to be considered a game. Note that none of the definitions of a "game" discussed in Chapter 3 include narrative as a fundamental component. In fact, *Foldit*, what I consider an exemplar of a successful serious game[4], has no embedded narrative (Jenkins, 2003). While the *ESP Game* does have a subtle framing narrative (of "reading someone's mind"), this narrative is in no way dominant, and was even dropped without detriment when the system was rebranded as the Google Image Labeler. This is not to say that narratives are superfluous in games. Narratives are extremely common in games, and are often used to provide motivation to play through the use of plot hooks (planted questions that players feel compelled to answer) and emotional proximity (empathy between the player and the characters or situations in the game) (Dickey, 2006). Hence narratives are highly important for many games: as mentioned previously, without the narrative it would be difficult to consider *Causlings* as a game—and even with the narrative, the lack of feedback makes *Causlings* a less than successful game. Interestingly, even Conway's *Game of Life* (1970)—which although called such, isn't normally considered a game due to its lack of interactivity—*does* has notable narrative framing, with certain cells "dying" due to "underpopulation" or "overcrowding" and others cells appearing as their neighbors "breed". In this way, narrativization is neither necessary or sufficient for a system to be considered a game or even a successful game.

Furthermore, as detailed above, certain forms of narrativization and game narrativization in particular may be at odds with systems thinking. For example, linear narratives may be tightly related with linear thinking, and thus less capable at supporting systems thinking. Indeed, framing a game's narrative as a linear sequence of steps (e.g., "go here, then here, then here") may reduce the opportunity for the player to engage with the complexity of the underlying game system, exploring any emergent behaviors that arise from the game's rules. Such engagement and exploration exercises a form of systems thinking; supporting this exploration is part of what makes games good at teaching systems thinking. Thus when a strong embedded narrative detracts from the exploration of emergent behaviors (and precludes *emergent gameplay*; Juul, 2005), it is interfering with a potentially unintentional form of systems thinking training. This phenomenon may be one of the

[4]Note that *Foldit* and other human computation games can often be considered as "puzzles"—that is, systems with a correct answer or outcome (Salen and Zimmerman, 2003). Although some scholars have argued that puzzles are not "true" games (e.g., Crawford, 1984), for the purposes of this dissertation I continue to consider them games.

reasons that games with highly linear narratives in which the player has no choice of actions (in which the player is "railroaded") are often considered to be less enjoyable—at the most extreme end, such linear narratives stop being games and simply become multimedia stories such as movies or books. In this way, the narrative that supports engagement with games can also interfere with the systems thinking-reliant activities that make games into fun learning engines.

Based on this consideration, explicit narrativization in games may be less beneficial or even harmful for games whose purpose is to develop systems thinking skills. Narrativization as a form of gamification (such as in the design of *Causlings*) does not necessarily ensure that a game will be engaging or that the system will achieve its non-entertainment goals. This research demonstrates that even narratives explicitly designed to support systems thinking may get in the way of such approaches. *Causlings* was explicitly designed to try and encourage (or at least not hinder) non-linear, web-like map building and systems thinking by casting the game as being about exploration and "spreading out" rather than a linear journey. Yet the narrative complexity of the system may have reduced the presence of systems thinking by introducing *too much* complexity to the already confusing causal mapping technique: in my interpretation of the results, the complexity of the narrative-based gamification encourages players to fall back on easier but less appropriate reductionist methods.

While this research demonstrates how the use of narrative may interfere with systems thinking, narrativization is just one of many gamification techniques. Educators and researchers need to continue to study the relationships between gamification and education—probing for what forms of gamification, while increasing engagement, may have costs in terms of increasingly important cognitive skills such as systems thinking. This dissertation this begins to point at some of the particular ways in which gamification may or may not be useful for education; however, further research is necessary.

Given that this research points out how certain forms of narrative may not be effective for encouraging systems thinking, how might educators best adopt game-like systems for supporting education regarding complex systems? I suggest that teachers in formal education contexts should focus on the use of more puzzle-like games that do not rely on strong embedded narratives. Puzzles and games that represent abstract systems with multiple interacting parts (and the resulting emergent effects) may be better suited for teaching systems thinking and improving systems thinking skills. For example, a game might be focused around the idea of "balancing" a large number of interrelated concepts, so that emergent and indirect effects could be explored. Such non-narrative puzzles may also be of potential use in assessing causal understanding. Moreover, a reduction in the reliance on linear narrative throughout the science curriculum (not only in game-based pedagogies) may be able to help support systems thinking. Much of formal science education—particularly for younger students—is structured around linear narratives (though often with a cyclical component): for example, students learn about the narrative of the water cycle or the narrative of a plant's life cycle. Indeed, such linear narratives may implicitly appear in the metaphors common to science education (Baumer, 2009): explaining a cell as a city may suggest a certain kind of "story" that frames the interactions of the cell in terms of linear causalities, hiding any emergent properties. Overall, this dissertation should suggests to educators in formal contexts how the common, narrative-centered form of serious game may encourage reductionist and linear thinking—other game formats are necessary for encouraging systems thinking.

Similarly, this dissertation also suggests that educators in informal education contexts (e.g., museums) should reduce their use of narrative-driven installations in order to support systems thinking. Such a change may be more difficult; as most museum installations are designed to be interacted with for a very short period of time, a driving linear narrative can be highly effective at framing the interaction and providing sufficient engagement. However, I suggest that that installations instead focus on providing compelling *emergent narratives* (Jenkins, 2003)—narratives that organically arise from the interaction with a game-like system. Indeed, installations might focus on creating compelling narratives of interaction: the story of what happened to the patron when they interacted with the exhibit, rather than a story told by the exhibit. Moreover, informal educational contexts may be uniquely suited to support systems thinking through multiplayer game systems, where emergent behaviors arise because of the unpredictability of multiple human actors. The development of such systems is important future work for enabling systems thinking education in informal contexts.

Overall, the results of this dissertation point at the difficulties of divorcing reductionist and linear thinking from gamified systems—particularly systems that are gamified using narrativization. While the contributions of this dissertation demonstrate techniques for assessing causal understanding, the development of further game systems that support system thinking (such as by forgoing embedded narratives) remains an open topic for educators and researchers.

7.5 Conclusion

In sum, the study presented in this dissertation demonstrates how causal map analysis enabled by the *Causlings* interactive system can be used to identify the presence of reductionist and linear thinking. Furthermore, this dissertation describes the characteristics of games that may make them less suitable for use as stand-alone assessments in traditional education contexts. I have detailed the key contributions of this dissertation, as well as some of the broader impacts in terms of the use of narrative in the development of game systems for supporting systems thinking. The next chapter describes further explorations and extensions to this research project.

Chapter 8. Future Work

This dissertation presents an initial exploration of how interactive computer systems and games can support assessing causal understanding. This work represents a preliminary study—there are many avenues for future work and extensions to the presented research. In this chapter I describe this future work, which includes: addressing current limitations of the presented prototype and study; integrating the *Causlings* prototype with the Causality Project (Tomlinson and Black, 2011a,b); and using interactive systems and games to crowdsource knowledge and measurements of causal relationships.

8.1 Addressing Study Limitations

The system and study presented in this dissertation have been subject to a number of limitations that can be addressed in future work. The interactive system is just a prototype, and so can benefit from further refinements. As dictated by the principles of HCI, the design and presentation of the interactive system, influenced and potentially biased the system's assessment of understanding. For example, the results show that players were more likely to include islands that were presented first (ordered alphabetically), and that players were remarkably more likely to build increasing bridges—bridges with the default influence direction. These effects were reminiscent of ordering effects found in multiple choice exam questions, as well as the potentially deleterious effects of "negative" choice options (e.g., Haladyna et al., 2002). Randomizing the order of presented elements in some way may reduce this bias (e.g., listing the islands in *Causlings* in a random order, rather than alphabetically), though it may also make the interface more difficult to comprehend and understand—particularly when the interface includes a large number of elements as in *Causlings*.

The interactive system could also be further extended to better study some of the behaviors demonstrated through the initial prototype. For example, detection of reductionist thinking as indicated by players' use of categorization can be automated through graph clustering methods. The confusion introduced by the complexity of the game's narrative framing may also be able to be addressed through carefully design tutorial levels (see Andersen et al., 2012) or other revised instructions. Finally, the influence of game dynamics could be further explored through the testing of a greater variety of game patterns (in addition to the Time Limit patterns considered in this study). Analysis of the effects of further design iterations—as well as more in-depth experiments as to the presence of reductionist and linear thinking (as might be identified through think-aloud methodologies, for example) may reveal more details and guidelines for how interactive systems and games in particular could be used to support assessment.

This study was also limited by the fact that it was not able to measure how accurately the *Causlings* system measured a player's true mental understanding on causality. The interactive system presented here relied instead on the established cognitive causal mapping technique to provide a valid assessment of understanding. Although cognitive causal mapping has been considered and used in prior research upon which this dissertation is based (i.e., Plate, 2006, 2010), the question of whether this assessment reflects a measurement of what is truly in a player's head remains open (see Chapter 2 for further discussion of the limitations of cognitive causal mapping).

Moreover, the study is unable to answer the question of whether the *Causlings* interactive system prototype provides a score that is reflective of a player's understanding of the complex causal system, or simply reflective of their understanding or ability to play the game. While there are some indications that the game prototype provides equivalent assessments to non-game based assessment methods (and indeed, the basic game form is very similar to non-game based methods), this validity remains to be proven. As discussed in Chapter 7, the complexity of the game framing may have interfered with the assessment performed by the interactive system. My consideration of the process by which people construct causal maps relies on an interpretation of game actions: what islands were used and what bridges were built. Although these actions reflect the construction process, they offer no insights into how players considered their own map building while playing the game—this study includes post-game interview responses, but no player response data gather *in situ*. It is quite possible that my analysis missed certain patterns or common thought processes that were not represented by game data, and it is unknown if discovered trends had any conscious basis. The exploration of how people cognitively reify causal maps would be greatly supported by further study into the intent of player's moves—such as could be gathered through a think-aloud study procedure. Similarly, the reliability of this assessment—whether it produces similar results over time—also remains to be demonstrated.

Addressing these limitations can help strengthen the arguments made in this dissertation, as well as further our understanding of how interactive systems and game dynamics may influence assessment of causal understanding—or indeed if such systems can even be used regularly.

8.2 Integration with The Causality Project

Described previously in Chapter 2, the Causality Project is an online, user-contributed database and visualization tool for the web of causal relationships between factors in the domain of environmental sustainability. The *Causlings* prototype was developed as an extension of the Causality Project's codebase, and the prototype's formulation of causality was significantly influenced by this project. Indeed, the interactive system's modeling of causal relationships has an almost direct correspondence with the broader project, with *Causlings* islands being the equivalent of Causality's "issues" (i.e., causal factors), and bridges being the equivalent of Causality's "relationships" (i.e., causal relationships)[1]. Because of these equivalences, as well as sharing the same codebase, *Causlings* could easily be integrated with the Causality Project in order to enable the assessment of a wider variety of complex systems. The Causality Project includes causal relationships among and within a great number of sustainability-related complex systems (a variety that increases as the database grows)—systems that can be isolated by selecting a subset of causal factors from the Causality database. *Causlings* could thus choose a collection of related causal factors (islands) for players to map, thereby assessing their understanding of the causal relationships in the complex system made up of those islands. In terms of implementation, rather than use a pre-defined (and hard-coded) set of islands, the interactive system could simply choose a set of connected issues from the Causality database and present those to the player—choosing islands to use dynamically rather than statically.

The primary challenge in this implementation is identifying the "true" causal relationships against which to measure players' performances and give them a score reflecting an assessment of their understanding. The current implementation of *Causlings* (as well as in Plate's (2006; 2010) research upon which my work is based) uses a causal map specifically generated and validated by a group of experts to act as the basis for "truth" of the complex system (see Chapter 4). Yet the Causality Project does not necessarily have a validated expert graph that can be considered the "truth" about all the complex systems that one may wish to assess with the *Causlings* game. Although in its current form the Causality Project's initial database of causal relationships has primarily been populated and constructed by expert users (e.g., researchers in the field of environmental science), the maps in this database have no guarantee of being either correct or complete—features that would be needed for the Causality database itself to properly act as ground truth against which to assess player's causal understanding. Thus further research into rapidly generating a database of valid and complete causal relationships for the Causality Project, such as by automatically extracting causality (e.g., Khoo, 1995; Pechsiri and Piriyakul, 2010) from the text of scientific publications,

[1]However, the Causality Project's database also includes relationships of the form "Factor A is a type of Factor B," marking a relationship between factors that while not causal, allows for causal chains to be explicated among a wider range of factors.

may be needed to effectively provide a large database of complex causal systems which a system such as *Causlings* could use as a basis for assessment.

Overall, extending this research by integrating the *Causlings* prototype with the Causality Project would allow the interactive system to be able to assess player knowledge of a wider variety of systems, offering the potential to study different publics' collective understanding of the complex causalities involved in environmental sustainability. The research presented in this dissertation provides the groundwork for such wide-scale measurements.

8.3 Crowdsourcing Knowledge of Causal Relationships

Assessing understanding of a wide variety of causal systems using the cognitive causal mapping process discussed in this dissertation would require a large number of "true" causal maps against which to compare subjects' inputs. These maps may be able to be synthesized from existing electronic databases of human knowledge—from Wikipedia to ConceptNet (Liu and Singh, 2004) to climate models such as CESM (Drake et al., 2005)—thereby alleviating the need to have experts generate valid causal graphs on which to base assessments. Alternatively, the research presented in this dissertation may also be able to support documenting people's collective knowledge of the causal relationships of a wide variety of systems. Aggregating knowledge of causal relationships across a large number of people allows for a greater variety of views and areas of expertise, and can include causal factors that touch upon a range of personal experience beyond that of any individual. Indeed, combining different people's understanding of causality can help to overcome individual limits of understanding, resulting in a greater collective intelligence or "wisdom of crowds" (Lévy, 2001; Surowiecki, 2005); harnessing such collective intelligence is one of the main draws of crowdsourced knowledge systems such as Wikipedia. Collective causal maps are already used in the field of management studies in order to render maps of highly complex systems that are beyond the experience of individual managers or workers—for example, Scavarda et al. (2006)

present a methodology for synthesizing causal knowledge from a wide variety of experts into a single causal graph. However, this process requires significant researcher intervention in the form of data collection and analysis, as subject responses are manually coded into causal relationships. The interactive computational system presented in this dissertation offers a tool for potentially automating the crowdsourcing for such causal knowledge.

Using the *Causlings* prototype involves players documenting their causal understanding through the construction of causal maps. In the system's current design, this process is used for assessment by comparing player-generated maps to those created by experts—but what if we instead considered each player's map as representative of the "truth" of their understanding, from their point of view? The bridges that players build could collectively reveal where people believe causal relationships to exist—and more people (or at least people with more expertise) agreeing on a bridge may indicate a higher likelihood that the bridge represents an actual causal relationship. By systematically aggregating the results of a large number of games played by a large number of people, it may be possible for researchers to automatically construct a collective causal map that, following the crowd wisdom theory, has the potentially to be more inclusive and more accurate than any map made by individuals. Furthermore, by looking how frequently bridges are built, as well as when in the sequence of bridges they are added to individual maps, researchers may be able to elicit the perceived causal strength of a relationship (i.e., earlier-made bridges may be more prominent and therefore have a greater magnitude of causal strength). In this way, by aggregating and comparing the strengths of individual bridges to one another, it may be possible to calculate a numerical value for the relative causal strengths of each relationship. Adding such a parameter to the Causality Project's database of causality in environmental sustainability can enable the system to model more accurate and in-depth simulations and measurements of the effects of specific changes on individual causal factors, greatly supporting the study of environmental systems.

Nevertheless, such modeling is a lofty goal, and there are a number of barriers that would need to be studied and overcome to crowdsource knowledge of causal relationships effectively. For one, the *Causlings* interactive prototype would need to be significantly redesigned—by removing the concept of a single "correct" map against which a player's map can be assessed, the system would lose the ability to give the player a score if the same game framing was maintained. Addressing this change will require careful design work, particularly as the evaluative study demonstrated the engaging power of a score to beat—a different scoring system may need to be designed, or even an entirely different game framing for collecting causal knowledge. But more importantly, further research is needed into whether a collective, crowdsourced understanding of causality does in fact converge towards expert understanding or some other "truth" that can effectively support understanding of and actions towards environmental sustainability. In other words: does the wisdom of crowds phenomenon apply to causal knowledge, particularly when collected through a game-based system? As people have difficulty understanding causality and complex systems individually, it is unclear whether they would demonstrate enough fundamental knowledge to allow a collective representation to overcome these flaws through aggregation[2]. Moreover, it is uncertain whether a game-based interactive system would be able to collect the specialized knowledge (e.g., of the biology of how algae process CO_2) needed to understand the complex systems surrounding sustainability.

8.4 Conclusion

The *Causlings* system presented in this dissertation provides an interface for people to reify their causal understanding as causal maps in an engaging and efficient manner, effectively automating documentation of individuals' causal knowledge. This research thus lays the groundwork for a wide range of future work involving the assessment and use of causal maps of complex systems, including how individual understanding might be aggregated to ultimately support greater environmental sustainability.

[2]Indeed, I hypothesize that collective causal maps would have the same flaws in causal understanding, such as the prominence of linear thinking and misunderstanding non-linear relationships.

Chapter 9. Conclusion

In this dissertation, I have studied people's understanding of causality, and how computational tools and games in particularly may be used to assess that understanding. I began by presenting a simple theoretical framework for considering systems thinking and set of semantics that can be used to talk about and visualize causal relationships (Chapter 2). I then showed how games are increasingly being positioned as effective tools for education, but there is little consideration on how games may be used for assessment as a component of education (Chapter 3). In response to this gap in the research literature, I presented *Causlings*, a prototype of an online interactive system in the form of a game that assesses players' understanding of causal relationships in complex systems through a previously validated cognitive mapping methodology (Chapter 4). I evaluated this prototype through a user study, considering both the structure and construction sequence of maps that were built through this system, as well as how the game framing may have influenced the reification of these maps (Chapter 5). Through this user study, I found that players had a strong tendency to categorize their maps, as well as to construct maps in a depth-first manner; in addition, the interactive system's game framing added levels of confusion to the causal mapping process—though the score presented at the end of the game was effective at motivating repeated play. (Chapter 6). Finally, I interrogated how these results demonstrate that the analysis enabled by the interactive system is able to detect reductionist and linear thinking, and how the requirements of learnability and feedback make games less effective for use in stand-alone assessments (Chapter 7), followed by a discussion of future extensions to this work (Chapter 8).

The interactive system presented and detailed in this dissertation improves upon existing methods for assessing causal understanding along a number of dimensions:

1. **Speed:** The computational system supporting the *Causlings* game drastically increased the speed at which games could be scored and causal understanding assessed. The accuracy of a causal map could be determined immediately, allowing players to receive their scores and then play multiple games if desired. Indeed, this increased efficiency can allow the game to be used for real-time formative assessment as part of a broader learning process, rather than being limited to an end-of-term-style accountability assessment as with previous methods. Moreover, this automatic scoring can be extended to measures beyond expert agreement, such as the "web-like-ness" of the constructed map, if desired.
2. **Assessment Quality:** Moreover, the automatic collection of additional map construction data (such as spatial layout and construction sequence) also enables the interactive system to provide a more detailed assessment of causal understanding. The analysis used in this dissertation demonstrates how the structure of maps can be used to reveal reductionist thinking, while the sequence in which maps are built can reveal the degree of linear thinking. These components can add further depth to an assessment of causal understanding, beyond the simple measures of accuracy supported by the CMAST framework. In addition, the system can also automatically identify events such as when the player hesitates about a bridge or otherwise seem to change their conceptualization of the causal map (such by altering or deleting bridges) that may be used to identify particular flaws in understanding or instances of learning. While determining further forms of data that can be used to assess causal understanding most effectively is beyond the scope of this work, the presented system allows a greater variety of data to be collected, enabling more accurate and higher quality assessments.
3. **Scope:** Because of the automated scoring and increased speed, *Causlings* can be used to assess a larger numbers of subjects without increased effort on the part of the assessor. Assessed subjects can be distributed over time and space—such as how Turkers participated in the user study—and can even self-assess without the conscious guidance of an instructor

(though, as discussed above, the instructions for using the interactive system should be clarified in future iterations). In addition, this system has the capability to be easily expanded to include assessments of more topics and a wider variety of complex causal systems, described in Chapter 8.

4. **Participant Engagement:** Finally, the game framing of the interactive system offers a potentially more engaging method for assessing causal understanding—one that can be more enjoyable and interesting for the subject. However, this claim remains to be validated; I have no baseline against which to consider the measured engagement levels of the *Causlings* prototype. Nevertheless, the frequent highly positive responses to the particularly game-like elements, such as the game narrative and the interactive scoring—suggest that this assessment may on the whole be enjoyable, assuming the game framing can be centered with the assessment context in which the interactive system is used.

Overall, this dissertation demonstrates that the implemented *Causlings* system prototype is a tool that can be used to rapidly assess causal understanding of complex causal systems, while also providing a basis and example for the development for future techniques for measuring causal understanding and systems thinking, based on a wider variety of tracked data.

Furthermore, the study presented in this dissertation also offers further insights into the consideration of how people construct causal maps and how they understand the causal relationships of complex systems overall. For example, as detailed above, this study highlights the importance of categorizing causal factors for considering large complex systems, potentially as a strategy for bounding the system into a form that can be more readily intuited. Yet at the same time, this categorization suggests the prevalence of reductionist thinking, rather than systems thinking. This dissertation also offers a novel exploration of the sequence of map construction, demonstrating for example how an analysis of this process (i.e., detecting depth-first building sequences) may be used to help identify linear thinking—indeed, the results provide further support to previous research's recognition of the pervasiveness of linear thinking (e.g., Grotzer and Perkins, 2000; Plate, 2010;

Raia, 2005). While this study is primarily exploratory in nature, it does show how the process by which people construct causal maps—a topic that has not yet been explored in the literature—can significantly and informatively reflect causal understanding and interaction with complex causal systems.

In addition, I found that to some extent, game characteristics that are commonly used in gamification efforts—such as game narratives and providing "scores" as targets and feedback of success—are able to increase engagement with systems for assessing causal understanding. However, I also found that the complexity introduced by a game framing can interfere with the assessment process by forcing players to understanding one complex system (the game itself) in order to measure their understanding of another complex system. Beyond that, the characteristics of games as systems that give regular feedback and represent spaces in which players are "free to play" without fear of consequences stand in direct opposition to the requirements of traditional stand-alone assessments. This contrast suggests that games may be less effective in supporting assessments than they are in supporting other aspects of education such as learning.

With this dissertation, I offer three key novel and significant contributions to knowledge. First, I show that procedurally analyzing the structure of causal maps created in an interactive system might reveal the presence of reductionist thinking. The structure of a causal map can indicate the degree of categorization (e.g., the crispness of different node clusters); such categorization may be indicative of reductionist thinking as subjects try and reduce the complex system into less complex sub-systems that can then be recombined. By considering the level of categorization used in causal maps, we may be able to improved our assessment of a person's causal understanding in terms of his or her level of reductionist thinking.

Second, I demonstrate how the sequence in which a user constructs a causal map can be analyzed to detect the presence of linear thinking. Map construction sequence can be measured in terms of whether the map is built in a primarily depth-first or breadth-first manner. Depth-first building corresponds to a reliance on causal chains, which can be indicative of linear thinking. Moreover, this analysis can be automated by an interactive computer system, enabling the real-time assessment of a subject's use of linear thinking as opposed to systems thinking in the consideration of a complex causal system.

Third, I suggest that although games as a form of interactive system can be effective at supporting learning and education, they may be less readily effective in support stand-alone assessments—particularly in a traditional educational context. Playing a game (even a game for assessment) requires an *a priori* understanding of the complex system that underlies that game; in an assessment game, the results of the assessment may reflect the player's understanding of the game, rather than of the content being assessed. In this way, requiring players to master a complex game system in order to demonstrate their mastery and understanding of a different causal system can introduce significant biases that restrict the effectiveness of games for assessment. Moreover, the context of a traditional educational assessment precludes the ability of a game to give constant feedback (a feature necessary for a game to be both learnable and enjoyable). Without this feedback and the ability to freely experiment that it enabled, a game can rarely be successful. For these reasons, games may be less effective at supporting stand-alone assessments, though they still may be suitable for assessments embedded in larger learning contexts or systems.

Human understanding of the causal relationships in the social and natural world is sufficient for many day-to-day tasks, but deeply lacking for many others. People plan trips to the store in order to feed and provide for themselves, yet may not understand the broader effects their travel, their purchasing, and even their social acceptance of the whole procedure may cause in terms of the natural environment—and what's more, they are frequently not even aware of their misunderstandings. This dissertation offers a tool and analysis technique to aid in assessing what we don't know and in what ways we don't know it, as a starting point for learning how we can effectively take actions to begin mitigating the environmental damage we as humans have already unknowingly caused. But the environmental challenges facing our world is just one example of a complex system that is beyond the scope of our cognitive limits—international social economies, political and cultural infrastructures, and human health and behavior could all benefit from a clearer understanding of their constituent causal relationships and resultant emergent properties. This dissertation takes one small step towards better enabling the measurement this understanding and the determination of effective human actions for the betterment of society.

References

Aarseth, E. (2001). Computer game studies, year one. *Game Studies*, 1(1):1–15.

Abt, C. C. (1970). *Serious Games*. University Press of America.

Alcott, B. (2005). Jevons' paradox. *Ecological Economics*, 54(1):9–21. Available from: http://dx.doi.org/10.1016/j.ecolecon.2005.03.020.

Andersen, E., O'Rourke, E., Liu, Y.-E., Snider, R., Lowdermilk, J., Truong, D., Cooper, S., and Popovic, Z. (2012). The impact of tutorials on games of varying complexity. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, CHI '12, page 59–68, New York, NY, USA. ACM. Available from: http://doi.acm.org/10.1145/2207676.2207687.

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., and Wittrock, M. C. (2000). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives, Abridged Edition*. Allyn & Bacon.

Angelo, T. A. (1999). Doing assessment as if learning matters most. *AAHE Bulletin*, 51(9):3–6.

Antin, J. and Churchill, E. (2011). Badges in social media: A social psychological perspective. page 1–4, Vancouver, BC, Canada. Available from: http://gamification-research.org/wp-content/uploads/2011/04/03-Antin-Churchill.pdf.

Apperley, T. (2006). Genre and game studies: Toward a critical approach to video game genres. *Simulation & Gaming*, 37(1):6.

Assaraf, O. B. and Orion, N. (2010). System thinking skills at the elementary school level. *Journal of Research in Science Teaching*, 47(5):540–563. Available from: http://onlinelibrary.wiley.com/doi/10.1002/tea.20351/abstract.

Axelrod, R. M. (1976). *Structure of decision: The cognitive maps of political elites*. Princeton University Press.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. Worth Publishers.

Barber, W. and Badre, A. (1998). Culturability: The merging of culture and usability. In *Proceedings of the 4th Conference on Human Factors and the Web*, page 1–14.

Bartram, L. (1997). Perceptual and interpretative properties of motion for information visualization. In *Proceedings of the 1997 workshop on New paradigms in information visualization and manipulation*, NPIV '97, page 3–7, New York, NY, USA. ACM. Available from: http://doi.acm.org/10.1145/275519.275520.

Bateson, G. (1955). A theory of play and fantasy. *Psychiatric research reports*, 2(39):39–51.

Bateson, G. (1972). *Steps to an Ecology of Mind*. University of Chicago Press.

Baumer, E. (2009). *Computational metaphor identification to foster critical thinking and creativity*. Dissertation, University of California, Irvine. Available from: http://ericbaumer.com/publications/Baumer-Dissertation.pdf.

Bederson, B. B. and Quinn, A. J. (2011). Web workers unite! addressing challenges of online laborers. In *PART 1 ——— Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems*, CHI EA '11, page 97–106, New York, NY, USA. ACM. Available from: http://doi.acm.org/10.1145/1979602.1979606.

Begel, A. and Klopfer, E. (2007). Starlogo TNG: an introduction to game development. *Journal of E-Learning*. Available from: http://research.microsoft.com/en-us/um/people/abegel/papers/starlogo-tng.pdf.

Benkler, Y. (2002). Coase's penguin, or, linux and the nature of the firm. *Yale Law Journal*, 112(3):367–445.

Berryman, A. A. (1992). The orgins and evolution of predator-prey theory. *Ecology*, 73(5):1530–1535. ArticleType: research-article / Full publication date: Oct., 1992 / Copyright © 1992 Ecological Society of America. Available from: http://dx.doi.org/10.2307/1940005.

Björk, S. and Holopainen, J. (2005). *Patterns in game design*. Cengage Learning.

Bloom, B. S., Engelhart, M., Furst, E., Hill, W., and Krathwohl, D. R. (1956). *Taxonomy of educational objectives: the classification of educational goals; Handbook I: Cognitive Domain*. Longmans, Green, New York.

Bogost, I. (2007). *Persuasive games: The expressive power of videogames*. MIT Press.

Bogost, I. (2011). Persuasive games: exploitationware. *Gamasutra: The Art & Business of Making Games*. Available from: http://www.gamasutra.com/view/feature/6366/persuasive_games_exploitationware.php.

Boland, R. J. and Tenkasi, R. V. (1995). Perspective making and perspective taking in communities of knowing. *Organization science*, page 350–372.

Bonwell, C. C. and Eison, J. A. (1991). Active learning: Creating excitement in the classroom. ERIC digest. *ASHE-ERIC Higher Education Reports, The George Washington University*. Available from: http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED340272.

Booth Sweeney, L. and Sterman, J. D. (2000). Bathtub dynamics: initial results of a systems thinking inventory. *System Dynamics Review*, 16(4):249–286. Available from: http://onlinelibrary.wiley.com/doi/10.1002/sdr.198/abstract.

Born, M. (1949). *Natural philosophy of cause and chance*. Oxford.

Bougon, M. G. (1983). Uncovering cognitive maps: the self-q technique. In Morgan, G., editor, *Beyond Method: A Study of Organizational Research Strategies*, page 160–72. Sage, New York.

Brockmyer, J. H., Fox, C. M., Curtiss, K. A., McBroom, E., Burkhart, K. M., and Pidruzny, J. N. (2009). The development of the game engagement questionnaire: A measure of engagement in video game-playing. *Journal of Experimental Social Psychology*, 45(4):624–634. Available from: http://dx.doi.org/10.1016/j.jesp.2009.02.016.

Bruner, J. (1991). The narrative construction of reality. *Critical inquiry*, 18(1):1–21.

Bryant, S. L., Forte, A., and Bruckman, A. (2005). Becoming wikipedian: transformation of participation in a collaborative online encyclopedia. In *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*, pages 1–10, Sanibel Island, Florida, USA. ACM. Available from: http://portal.acm.org/citation.cfm?doid=1099203.1099205.

Brynjarsdottir, H., Håakansson, M., Pierce, J., Baumer, E., DiSalvo, C., and Sengers, P. (2012). Sustainably unpersuaded: how persuasion narrows our vision of sustainability. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, CHI '12, page 947–956, New York, NY, USA. ACM. Available from: http://doi.acm.org/10.1145/2208516.2208539.

Burgess, G. M., Clark, Thomas D., J., Hauser, Richard D., J., and Zmud, R. W. (1992). The application of causal maps to develop a collective understanding of complex organizational contexts in requirements analysis. *Accounting, Management and Information Technologies*, 2(3):143–164. Available from: http://dx.doi.org/10.1016/0959-8022(92)90013-I.

Caillois, R. (1961). *Man, play, and games*. Free Press.

Caporael, L. (1986). Anthropomorphism and mechanomorphism: two faces of the human machine. *Computers in human behavior*, 2(3):215–234.

Carley, K. and Palmquist, M. (1992). Extracting, representing, and analyzing mental models. *Social Forces*, 70(3):601–636. Available from: http://dx.doi.org/10.1093/sf/70.3.601.

Chaib-draa, B. (2002). Causal maps: theory, implementation, and practical applications in multiagent environments. *Knowledge and Data Engineering, IEEE Transactions on*, 14(6):1201 – 1217. Available from: http://dx.doi.org/10.1109/TKDE.2002.1047761.

Chen, C. (2005). Top 10 unsolved information visualization problems. *Computer Graphics and Applications, IEEE*, 25(4):12 – 16. Available from: http://dx.doi.org/10.1109/MCG.2005.91.

Chen, J. (2007). Flow in games (and everything else). *Commun. ACM*, 50(4):31–34. Available from: http://dx.doi.org/10.1145/1232743.1232769.

Cheung, G. and Huang, J. (2011). Starcraft from the stands: understanding the game spectator. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11, page 763–772, New York, NY, USA. ACM. Available from: http://doi.acm.org/10.1145/1978942.1979053.

Cialdini, R. B. (2001). *Influence: Science and practice*. Allyn and Bacon Boston.

Conway, J. (1970). The game of life. *Scientific American*, 223(4):4.

Crawford, C. (1984). *The Art of Computer Game Design: Reflections of a Master Game Designer*. McGraw-Hill Osborne.

Csikszentmihalyi, M. (1991). *Flow: The Psychology of Optimal Experience*. Harper Perennial.

Desurvire, H. and Wiberg, C. (2009). Game usability heuristics (PLAY) for evaluating and designing better games: The next iteration. In Ozok, A. and Zaphiris, P., editors, *Online Communities and Social Computing*, volume 5621 of *Lecture Notes in Computer Science*, pages 557–566. Springer Berlin / Heidelberg.

Deterding, S. (2011). Meaningful play. getting »Gamification« right. Available from: http://www.slideshare.net/dings/meaningful-play-getting-gamification-right.

Deterding, S., Dixon, D., Khaled, R., and Nacke, L. (2011a). From game design elements to gamefulness: defining "gamification". In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, MindTrek '11, page 9–15, New York, NY, USA. ACM. Available from: http://doi.acm.org/10.1145/2181037.2181040.

Deterding, S., Sicart, M., Nacke, L., O'Hara, K., and Dixon, D. (2011b). Gamification. using game-design elements in non-gaming contexts. In *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems*, CHI EA '11, page 2425–2428, New York, NY, USA. ACM. Available from: http://doi.acm.org/10.1145/1979482.1979575.

Dickey, M. D. (2006). Game design narrative for learning: Appropriating adventure game design narrative devices and techniques for the design of interactive learning environments. *Educational Technology Research and Development*, 54(3):245–263. ArticleType: research-article / Full publication date: Jun., 2006 / Copyright © 2006 Springer. Available from: http://www.jstor.org/stable/30221219.

Dillahunt, T., Mankoff, J., Paulos, E., and Fussell, S. (2009). It's not all about "Green". page 255. ACM Press. Available from: http://dx.doi.org/10.1145/1620545.1620583.

DML (2011). Badges for lifelong learning. Retrieved October 18, 2012, from http://www.dmlcompetition.net/.

Doucet, L. and Srinivasan, V. (2010). Designing entertaining educational games using procedural rhetoric. In *Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games - Sandbox '10*, pages 5–10, Los Angeles, California. Available from: http://dx.doi.org/10.1145/1836135.1836136.

Dourish, P. (2010). HCI and environmental sustainability: the politics of design and the design of politics. In *Proceedings of the 8th ACM Conference on Designing Interactive Systems*, DIS '10, page 1–10, New York, NY, USA. ACM. Available from: http://doi.acm.org/10.1145/1858171.1858173.

Downs, J. S., Holbrook, M. B., Sheng, S., and Cranor, L. F. (2010). Are your participants gaming the system?: screening mechanical turk workers. In *Proceedings of the 28th international conference on Human factors in computing systems*, CHI '10, page 2399–2402, New York, NY, USA. ACM. Available from: http://doi.acm.org/10.1145/1753326.1753688.

Doyle, J. K. (1997). The cognitive psychology of systems thinking. *System Dynamics Review*, 13(3):253–265. Available from: http://onlinelibrary.wiley.com/doi/10.1002/(SICI)1099-1727(199723)13:3%3C253::AID-SDR129%3E3.0.CO;2-H/abstract.

Drake, J. B., Jones, P. W., and Carr, G. R. (2005). Overview of the software design of the community climate system model. *International Journal of High Performance Computing Applications*, 19(3):177–186. Available from: http://dx.doi.org/10.1177/1094342005056094.

Ducheneaut, N., Yee, N., Nickell, E., and Moore, R. J. (2006). "Alone together?": exploring the social dynamics of massively multiplayer online games. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 407–416, Montréal, Québec, Canada. ACM. Available from: http://dx.doi.org/10.1145/1124772.1124834.

Eden, C. (1992). On the nature of cognitive maps. *Journal of Management Studies*, 29(3):261–265. Available from: http://onlinelibrary.wiley.com/doi/10.1111/j.1467-6486.1992.tb00664.x/abstract.

Eden, C., Ackermann, F., and Cropper, S. (1992). The analysis of cause maps. *Journal of Management Studies*, 29(3):309–324. Available from: http://onlinelibrary.wiley.com/doi/10.1111/j.1467-6486.1992.tb00667.x/abstract.

Elmqvist, N. (2004). *Visualization of Causal Relations*. Licentiate of philosophy thesis, Chalmers University of Technology and Göteborg University, Göteborg, Sweden.

Elmqvist, N. and Tsigas, P. (2004). Animated visualization of causal relations through growing 2D geometry. *Information Visualization*, 3(3):154–172. Available from: http://dx.doi.org/10.1057/palgrave.ivs.9500074.

EPA (2011). What you can do - drive wise. Retrieved April 18, 2012, from http://www.epa.gov/air/actions/drive_wise.html.

Fahey, L. and Narayanan, V. K. (1989). Linking changes in revealed causal maps and environmental change: An empirical study. *Journal of Management Studies*, 26(4):361–378.

Available from: http://onlinelibrary.wiley.com/doi/10.1111/j.1467-6486.1989.tb00734.x/abstract.

Falcon, A. (2011). Aristotle on causality. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Fall 2011 edition. Available from: http://plato.stanford.edu/archives/fall2011/entries/aristotle-causality/.

Federoff, M. A. (2002). *Heuristics and usability guidelines for the creation and evaluation of fun in video games*. Master's thesis, Indiana University.

Fitzherbert, E. B., Struebig, M. J., Morel, A., Danielsen, F., Brühl, C. A., Donald, P. F., and Phalan, B. (2008). How will oil palm expansion affect biodiversity? *Trends in Ecology & Evolution*, 23(10):538–545. Available from: http://dx.doi.org/10.1016/j.tree.2008.06.012.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist*, 34(10):906.

Fogg, B. J. (2003). *Persuasive technology: using computers to change what we think and do*. Morgan Kaufmann.

Forrester, J. W. (1961). *Industrial dynamics*. Cambridge, MIT Press.

Forster, E. M. (1927). *Aspects of the Novel*. Harcourt.

Frasca, G. (2001). *Videogames of the oppressed: Videogames as a means for critical thinking and debate*. PhD thesis, Georgia Institute of Technology.

Froehlich, J., Dillahunt, T., Klasnja, P., Mankoff, J., Consolvo, S., Harrison, B., and Landay, J. A. (2009). UbiGreen: investigating a mobile tool for tracking and supporting green transportation habits. In *Proceedings of the 27th international conference on Human factors in computing systems*, pages 1043–1052, Boston, MA, USA. ACM. Available from: http://dx.doi.org/10.1145/1518701.1518861.

Froehlich, J., Findlater, L., and Landay, J. (2010). The design of eco-feedback technology. In *Proceedings of the 28th international conference on Human factors in computing systems*, pages 1999–2008, Atlanta, Georgia, USA. ACM. Available from: http://portal.acm.org/citation.cfm?id=1753326.1753629&coll=GUIDE&dl=ACM&CFID=89970921&CFTOKEN=40393680.

Fu, F.-L., Su, R.-C., and Yu, S.-C. (2009). EGameFlow: a scale to measure learners' enjoyment of e-learning games. *Computers & Education*, 52(1):101–112. Available from: http://dx.doi.org/10.1016/j.compedu.2008.07.004.

Fujiki, Y., Kazakos, K., Puri, C., Buddharaju, P., Pavlidis, I., and Levine, J. (2008). NEAT-o-Games: blending physical activity and fun in the daily routine. *Comput. Entertain.*, 6(2):1–22. Available from: http://dx.doi.org/10.1145/1371216.1371224.

Fullerton, T., Swain, C., and Hoffman, S. (2008). *Game design workshop: a playcentric approach to creating innovative games*. Morgan Kaufmann.

Games, I. A. (2008). Three dialogs: A framework for the analysis and assessment of twenty-first-century literacy practices, and its use in the context of game design within "Gamestar mechanic". *E-Learning*, 5(4):396–417. Available from: http://dx.doi.org/10.2304/elea.2008.5.4.396.

"Games for Change" (n.d.). Games for change. Retieved October 18, 2012, from http://www.gamesforchange.org/.

Garris, R., Ahlers, R., and Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming*, 33(4):441–467. Available from: http://sag.sagepub.com/content/33/4/441.

Gee, J. P. (2003a). *What video games have to teach us about learning and literacy*. Macmillan.

Gee, J. P. (2003b). What video games have to teach us about learning and literacy. *Comput. Entertain.*, 1(1):20–20. Available from: http://doi.acm.org/10.1145/950566.950595.

Gee, J. P. (2004). *Situated Language And Learning: A Critique Of Traditional Schooling*. Psychology Press.

Gigerenzer, G. (2000). *Adaptive thinking: rationality in the real world*. Oxford University Press.

Gigerenzer, G. and Todd, P. M. (1999). *Simple heuristics that make us smart*. Oxford University Press, USA.

Gillenwater, M., Broekhoff, D., Trexler, M., Hyman, J., and Fowler, R. (2007). Policing the voluntary carbon market. *Nature Reports Climate Change*, (0711):85–87. Available from: http://dx.doi.org/10.1038/climate.2007.58.

Gipps, C. V. (1994). *Beyond Testing: Towards a Theory of Educational Assessment*. Psychology Press.

Gobert, J. and Clement, J. (1999). Effects of student-generated diagrams versus student-generated summaries on conceptual understanding of causal and dynamic knowledge in plate tectonics. *Journal of Research in Science Teaching*, 36(1):39–53. Available from: http://dx.doi.org/10.1002/(SICI)1098-2736(199901)36:1%3C39::AID-TEA4%3E3.0.CO;2-I.

Goffman, E. (1959). The presentation of self in everyday life. *Anchor Books*.

Goldberg, D., Nichols, D., Oki, B. M., and Terry, D. (1992). Using collaborative filtering to weave an information tapestry. *Commun. ACM*, 35(12):61–70. Available from: http://dx.doi.org/10.1145/138859.138867.

Goldstone, R. L. and Wilensky, U. (2008). Promoting transfer by grounding complex systems principles. *Journal of the Learning Sciences*, 17(4):465–516. Available from: http://www.tandfonline.com/doi/abs/10.1080/10508400802394898.

Goodwin, J. S. and Franklin, S. G. (1994). The beer distribution game: Using simulation to teach systems thinking. *Journal of Management Development*, 13(8):7–15. Available from: http://dx.doi.org/10.1108/02621719410071937.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L., Kushnir, T., and Danks, D. (2004). A theory of causal learning in children: causal maps and bayes nets. *Psychological Review; Psychological Review*, 111(1):3.

Gopnik, A. and Schulz, L. (2007). *Causal Learning: Psychology, Philosophy, And Computation*. Oxford University Press.

Grier, D. A. (2005). *When Computers Were Human*. Princeton University Press.

Grossman, T., Fitzmaurice, G., and Attar, R. (2009). A survey of software learnability: metrics, methodologies and guidelines. In *Proceedings of the 27th international conference on Human factors in computing systems*, CHI '09, page 649–658, New York, NY, USA. ACM. Available from: http://doi.acm.org/10.1145/1518701.1518803.

Grotzer, T. A. and Perkins, D. N. (2000). A taxonomy of causal models: The conceptual leaps between models and students' reflections on them. In *Annual meeting of the National Association for Research in Science Teaching*, New Orleans.

Grotzer, T. A. and Sudbury, M. (2000). Moving beyond underlying linear causal models of electrical circuits. In *Annual meeting of the National Association of Research in Science Teaching*, New Orleans, LA.

Gustafsson, A., Katzeff, C., and Bang, M. (2009). Evaluation of a pervasive game for domestic energy engagement among teenagers. *Comput. Entertain.*, 7(4):1–19. Available from: http://dx.doi.org/10.1145/1658866.1658873.

Haladyna, T. M., Downing, S. M., and Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3):309–333. Available from: http://www.tandfonline.com/doi/abs/10.1207/S15324818AME1503_5.

Halter, E. (2006). *From Sun Tzu to Xbox: War And Video Games*. PublicAffairs.

Hamari, J. and Eranti, V. (2011). Framework for designing and evaluating game achievements. In *Think Design Play: The fifth international conference of the Digital Research Association (DIGRA)*, page 20, Hilversum, the Netherlands: DiGRA/Utrecht School of the Arts.

Hayles, N. K. (1999). *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. University Of Chicago Press.

He, H. A., Greenberg, S., and Huang, E. M. (2010). One size does not fit all: applying the transtheoretical model to energy feedback technology design. In *Proceedings of the 28th international conference on Human factors in computing systems*, pages 927–936, Atlanta, Georgia, USA. ACM. Available from: http://dx.doi.org/10.1145/1753326.1753464.

Heimerl, K., Gawalt, B., Chen, K., Parikh, T., and Hartmann, B. (2012). CommunitySourcing: engaging local crowds to perform expert work via physical kiosks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1539–1548, Austin, TX, USA. ACM. Available from: http://doi.acm.org/10.1145/2207676.2208619.

Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research*, 58(1):47–77. Available from: http://dx.doi.org/10.3102/00346543058001047.

Herring, H. (2008). Rebound effect. Retrieved July 17, 2012, from http://www.eoearth.org/article/Rebound_effect.

Hitchcock, C. (2001). A tale of two effects. *The Philosophical Review*, 110(3):361–396.

Hitchcock, C. (2007). On the importance of causal taxonomy. In *Causal learning: Psychology, philosophy and computation*, page 101–114.

Hmelo-Silver, C. E. and Azevedo, R. (2006). Understanding complex systems: Some core challenges. *Journal of the Learning Sciences*, 15(1):53–61. Available from: http://www.tandfonline.com/doi/abs/10.1207/s15327809jls1501_7.

Holland, J. H. (1999). *Emergence: From chaos to order*. Basic Books.

Hopkins, P. L. (1992). Simulating hamlet in the classroom. *System Dynamics Review*, 8(1):91–98. Available from: http://onlinelibrary.wiley.com/doi/10.1002/sdr.4260080109/abstract.

Howe, J. (2008). *Crowdsourcing: How the power of the crowd is driving the future of business*. Century.

Huizinga, J. (1955). *Homo ludens: A study of the play-element in culture*. Beacon.

Hulswit, M. (2004). A short history of 'Causation'. *SEED Journal (Semiotics, Evolution, Energy, and Development)*, 4(3).

Hume, D. (1748). *An Enquiry Concerning Human Understanding*. Oxford University Press Edition (2000).

Hung, A. C. Y. (2011). *The Work of Play: Meaning-Making in Video Games*. Peter Lang Publishing.

Hunicke, R. and Chapman, V. (2004). AI for dynamic difficulty adjustment in games. In *Challenges in Game Artificial Intelligence AAAI Workshop*, page 91–96.

Hutchins, B. (2008). Signs of meta-change in second modernity: The growth of e-sport and the world cyber games. *New Media & Society*, 10(6):851–869. Available from: http://dx.doi.org/10.1177/1461444808096248.

IPCC (2007). Summary for policymakers. *Climate Change 2007: The Physical Science Basis*.

Ipeirotis, P. G. (2010). Demographics of mechanical turk. working paper, New York University.

Ito, M. (2010). *Hanging Out, Messing Around, and Geeking Out: Kids Living and Learning With New Media*. MIT Press.

Jacobs, F. R. (2000). Playing the beer distribution game over the internet. *Production and Operations Management*, 9(1):31–39. Available from: http://onlinelibrary.wiley.com/doi/10.1111/j.1937-5956.2000.tb00321.x/abstract.

Jacobson, M. J. and Wilensky, U. (2006). Complex systems in education: Scientific and educational importance and implications for the learning sciences. *Journal of the Learning Sciences*, 15(1):11–34. Available from: http://www.tandfonline.com/doi/abs/10.1207/s15327809jls1501_4.

Jakobsson, M. (2007). Playing with the rules: Social and cultural aspects of game rules in a console game club. In *Situated Play, Proceedings of the Digital Games Research Association (DiGRA) Conference*.

Jakobsson, M. (2011). The achievement machine: understanding xbox 360 achievements in gaming practices. *Game Studies*, 11(1).

Jenkins, H. (2003). Game design as narrative architecture. *First person: New media as story, performance, and game*.

Jenkins, H. (2006). *Fans, Bloggers, and Gamers: Exploring Participatory Culture*. NYU Press.

Jenkins, H. (2009). *Confronting the Challenges of Participatory Culture: Media Education for the 21st Century*. MIT Press.

Jenkins, M. and Johnson, G. (1997). Linking managerial cognition and organizational performance: A preliminary investigation using causal maps. *British Journal of Management*, 8:77–90. Available from: http://onlinelibrary.wiley.com/doi/10.1111/1467-8551.8.s1.7/abstract.

Juul, J. (2003). The game, the player, the world: Looking for a heart of gameness. In *Level Up: Digital Games Research Conference Proceedings*, Utrecht: Utrecht University.

Juul, J. (2005). *Half-real: Video games between real rules and fictional worlds*. MIT Press.

Kadaba, N. R., Irani, P. P., and Leboe, J. (2007). Visualizing causal semantics using animations. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1254 –1261. Available from: http://dx.doi.org/10.1109/TVCG.2007.70528.

Kadaba, N. R., Irani, P. P., and Leboe, J. (2009). Analyzing animated representations of complex causal semantics. In *Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization*, APGV '09, page 77–84, New York, NY, USA. ACM. Available from: http://doi.acm.org/10.1145/1620993.1621009.

Kato, P. M., Cole, S. W., Bradlyn, A. S., and Pollock, B. H. (2008). A video game improves behavioral outcomes in adolescents and young adults with cancer: A randomized trial. *Pediatrics*, 122(2):e305–e317. Available from: http://pediatrics.aappublications.org/content/122/2/e305.

Kearney, A. R., Bradley, G., Kaplan, R., and Kaplan, S. (1999). Stakeholder perspectives on appropriate forest management in the pacific northwest. *Forest Science*, 45(1):62–73.

Kearney, A. R. and Kaplan, S. (1997). Toward a methodology for the measurement of knowledge structures of ordinary people the conceptual content cognitive map (3CM). *Environment and Behavior*, 29(5):579–617. Available from: `http://dx.doi.org/10.1177/001391659729500l`.

Khatib, F., Cooper, S., Tyka, M. D., Xu, K., Makedon, I., Popović, Z., Baker, D., and Players, F. (2011a). Algorithm discovery by protein folding game players. *Proceedings of the National Academy of Sciences*, 108(47):18949–18953. Available from: `http://www.pnas.org/content/108/47/18949`.

Khatib, F., DiMaio, F., Group, F. C., Group, F. V. C., Cooper, S., Kazmierczyk, M., Gilski, M., Krzywda, S., Zabranska, H., Pichova, I., Thompson, J., Popović, Z., Jaskolski, M., and Baker, D. (2011b). Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nature Structural & Molecular Biology*, 18(10):1175–1177. Available from: `http://dx.doi.org/10.1038/nsmb.2119`.

Khoo, C. S. (1995). *Automatic identification of causal relations in text and their use for improving precision in information retrieval*. Dissertation, School of Information Studies of Syracuse University, Syracuse, NY.

Kittur, A., Chi, E. H., and Suh, B. (2008). Crowdsourcing user studies with mechanical turk. In *Proc. CHI 2008*, pages 453–456, Florence, Italy. ACM. Available from: `http://dx.doi.org/10.1145/1357054.1357127`.

Klopfer, E., Osterweil, S., and Salen, K. (2009a). Moving learning games forward. Technical report, Education Arcade, MIT.

Klopfer, E., Roque, R., Huang, W., Wendel, D., and Scheintaub, H. (2009b). The simulation cycle: combining games, simulations, engineering and science using StarLogo TNG. *E-Learning and Digital Media*, 6(1):71. Available from: `http://dx.doi.org/10.2304/elea.2009.6.1.71`.

Kobourov, S. G. (2007). Force-directed drawing algorithms. In Tamassia, R., editor, *Handbook of Graph Drawing and Visualization (Discrete Mathematics and Its Applications)*. Chapman & Hall/CRC.

Kosko, B. (1986). Fuzzy cognitive maps. *International Journal of Man-Machine Studies*, 24(1):65–75. Available from: `http://dx.doi.org/10.1016/S0020-7373(86)80040-2`.

Koster, R. (2005). *A Theory Of Fun For Game Design*. Paraglyph Press.

Kuznetsov, S. (2006). Motivations of contributors to wikipedia. *SIGCAS Comput. Soc.*, 36(2):1. Available from: `http://dx.doi.org/10.1145/1215942.1215943`.

Lave, J. and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press.

Lee, C.-S., Wang, M.-H., Chaslot, G., Hoock, J.-B., Rimmel, A., Teytaud, O., Tsai, S.-R., Hsu, S.-C., and Hong, T.-P. (2009). The computational intelligence of MoGo revealed in taiwan's computer go tournaments. *Computational Intelligence and AI in Games, IEEE Transactions on*, 1(1):73 –89. Available from: http://dx.doi.org/10.1109/TCIAIG.2009.2018703.

Lévy, P. (2001). Collective intelligence. *Reading digital culture*, 4:253.

Li, I., Dey, A., and Forlizzi, J. (2010). A stage-based model of personal informatics systems. In *Proceedings of the 28th international conference on Human factors in computing systems*, pages 557–566, Atlanta, Georgia, USA. ACM. Available from: http://portal.acm.org/citation.cfm?id=1753326.1753409&coll=Portal&dl=ACM&CFID=89970548&CFTOKEN=50636749.

Light, J. S. (1999). When computers were women. *Technology and Culture*, 40(3):455–483. ArticleType: research-article / Full publication date: Jul., 1999 / Copyright © 1999 Society for the History of Technology.

Lin, J., Mamykina, L., Lindtner, S., Delajoux, G., and Strub, H. (2006). Fish'n'Steps: encouraging physical activity with an interactive computer game. In *UbiComp 2006: Ubiquitous Computing*, pages 261–278. Available from: http://dx.doi.org/10.1007/11853565_16.

Lindley, C. (2005). Game space design foundations for trans-reality games. In *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology*, pages 397–404. ACM. Available from: http://dx.doi.org/10.1145/1178477.1178569.

Linkie, M., Martyr, D. J., Holden, J., Yanuar, A., Hartana, A. T., Sugardjito, J., and Leader-Williams, N. (2003). Habitat destruction and poaching threaten the sumatran tiger in kerinci seblat national park, sumatra. *Oryx*, 37(1):41–48.

Liu, H. and Singh, P. (2004). ConceptNet — a practical commonsense reasoning tool-kit. *BT Technology Journal*, 22(4):211–226. Available from: http://dx.doi.org/10.1023/B:BTTJ.0000047600.45421.6d.

Lowood, H. (2007). Found technology: Players as innovators in the making of machinima. In *Digital Youth, Innovation, and the Unexpected*, The John D. and Catherine T. MacArthur Foundation Series on Digital Media and Learning, page 165–196. MIT Press.

Lyons, D. E., Long, J., Goraya, R., Lu, J., and Tomlinson, B. (2012). Cultivating environmental systems thinking with karunatree. Technical Report LUCI 2012-002. Retrieved October 18, 2012, from http://luci.ics.uci.edu/#techreports.

Lyons, D. E. and Santos, L. R. (2006). Ecology, domain specificity, and the origins of theory of mind: Is competition the catalyst? *Philosophy Compass*, 1(5):481–492. Available from: http://onlinelibrary.wiley.com/doi/10.1111/j.1747-9991.2006.00032.x/full.

Lyons, D. E. and Tomlinson, B. (2010). Designing games that cultivate environmental literacy. Presented at the Games+Learning+Society Conference, Madison, WI.

Mackie, J. L. (1980). *The cement of the universe*. Clarendon Press Oxford.

Markóczy, L. and Goldberg, J. (1995). A method for eliciting and comparing causal maps. *Journal of Management*, 21(2):305–333. Available from: http://dx.doi.org/10.1177/014920639502100207.

Marshall, P. (2007). Do tangible interfaces enhance learning? In *Proceedings of the 1st international conference on Tangible and embedded interaction*, pages 163–170. ACM, ACM Press. Available from: http://dx.doi.org/10.1145/1226969.1227004.

Martin, J. H. (1990). A new iron age, or a ferric fantasy. *US JGOFS Newsletter*, 1(4):5–6.

Mayr, O. (1986). *Authority, liberty, & automatic machinery in early modern Europe*. Johns Hopkins University Press Baltimore.

McClure, J. R., Sonak, B., and Suen, H. K. (1999). Concept map assessment of classroom learning: Reliability, validity, and logistical practicality. *Journal of research in science teaching*, 36(4):475–492.

McGonigal, J. (2007). Why i love bees: A case study in collective intelligence gaming. *The John D. and Catherine T. MacArthur Foundation Series on Digital Media and Learning*, pages 199–227.

McGonigal, J. (2011). *Reality is Broken: Why Games Make Us Better and How They Can Change the World*. Penguin.

Menzies, P. (2009). Counterfactual theories of causation. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Fall 2009 edition. Available from: http://plato.stanford.edu/archives/fall2009/entries/causation-counterfactual/.

Mesarović, M. D. and Takahara, Y. (1975). *General Systems Theory: Mathematical Foundations*. Academic Press.

Michotte, A. (1963). *The perception of causality*. Oxford, England.

Montola, M., Stenros, J., and Waern, A. (2009). *Pervasive Games: Theory and Design*. Morgan Kaufmann Publishers Inc.

Moxnes, E. (2000). Not only the tragedy of the commons: misperceptions of feedback and policies for sustainable development. *System Dynamics Review*, 16(4):325–348. Available from: http://onlinelibrary.wiley.com/doi/10.1002/sdr.201/abstract.

Mueller, F. F., Edge, D., Vetere, F., Gibbs, M. R., Agamanolis, S., Bongers, B., and Sheridan, J. G. (2011). Designing sports: a framework for exertion games. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11, page 2651–2660, New York, NY, USA. ACM. Available from: http://doi.acm.org/10.1145/1978942.1979330.

Murray, J. H. (1997). *Hamlet on the holodeck: the future of narrative in cyberspace*. Simon and Schuster.

Nadkarni, S. and Narayanan, V. K. (2005). Validity of the structural properties of text-based causal maps: An empirical assessment. *Organizational Research Methods*, 8(1):9–40. Available from: http://orm.sagepub.com/content/8/1/9.

National Research Council (1996). *National Science Education Standards*. Retrieved April 04, 2012, from http://www.nap.edu/openbook.php?record_id=4962.

NDP Group (2011). Digital gamers closing on core gamers for top gamer segment. Retrieved April 04, 2012, from https://www.npd.com/press/releases/press_110627a.html.

Nicolini, D. (1999). Comparing methods for mapping organizational cognition. *Organization Studies*, 20(5):833–860. Available from: http://dx.doi.org/10.1177/0170840699205006.

Nienhaus, J., Ziegenbein, A., and Schoensleben, P. (2006). How human behaviour amplifies the bullwhip effect. a study based on the beer distribution game online. *Production Planning & Control*, 17(6):547–557. Available from: http://www.tandfonline.com/doi/abs/10.1080/09537280600866587.

Novak, J. D. (2010). *Learning, Creating, and Using Knowledge: Concept Maps as Facilitative Tools in Schools and Corporations*. Taylor & Francis.

Novak, J. D. and Cañas, A. J. (2008). The theory underlying concept maps and how to construct and use them. *Florida Institute for Human and Machine Cognition*, Technical Report IHMC CmapTools 2006-01 Rev 2008-01.

Odum, H. T. (1983). *Systems Ecology; an introduction*. Wiley (New York).

Özesmi, U. and Özesmi, S. L. (2003). A participatory approach to ecosystem conservation: Fuzzy cognitive maps and stakeholder group analysis in uluabat lake, turkey. *Environmental Management*, 31(4):0518–0531. Available from: http://dx.doi.org/10.1007/s00267-002-2841-1.

Özesmi, U. and Özesmi, S. L. (2004). Ecological models based on people's knowledge: a multistep fuzzy cognitive mapping approach. *Ecological Modelling*, 176(1–2):43–64. Available from: http://dx.doi.org/10.1016/j.ecolmodel.2003.10.027.

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. Basic Books, Inc.

Parker, F. (2008). The significance of jeep tag: On player-imposed rules in video games. *Loading... The Journal of the Canadian Game Studies Association*, 2(3).

Parlett, D. (1999). *The Oxford history of board games*. Oxford University Press, New York.

Partnership for 21st Century Skills (2001). Framework for 21st century learning. Retrieved April 04, 2012, from http://www.p21.org/overview/skills-framework/260.

Pearce, C. (2006). Productive play: Game culture from the bottom up. *Games and Culture*, 1(1):17–24. Available from: http://dx.doi.org/10.1177/1555412005281418.

Pearce, C., Fullerton, T., Fron, J., and Morie, J. F. (2007). Sustainable play: Toward a new games movement for the digital age. *Games and Culture*, 2(3):261.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.

Pearl, J. (2001). Direct and indirect effects. In *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, UAI'01, page 411–420, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc. Available from: http://dl.acm.org/citation.cfm?id=2074022.2074073.

Pechsiri, C. and Piriyakul, R. (2010). Explanation knowledge graph construction through causality extraction from texts. *Journal of Computer Science and Technology*, 25(5):1055–1070. Available from: http://dx.doi.org/10.1007/s11390-010-9387-0.

Penner, D. E. (2000). Explaining systems: Investigating middle school students' understanding of emergent phenomena. *Journal of Research in Science Teaching*, 37(8):784–806. Available from: http://onlinelibrary.wiley.com/doi/10.1002/1098-2736(200010)37:8%3C784::AID-TEA3%3E3.0.CO;2-E/abstract.

Pew Center on Global Climate Change (2009). Key scientific developments since the IPCC fourth assessment report. Technical report. Retrieved April 05, 2012, from http://www.pewclimate.org/brief/science-developments/June2009.

Piaget, J. (1930). *The Child's Conception of Physical Causality*. K Paul, London.

Plate, R. (2006). *Assessing the Effectiveness of Systems-Oriented Instruction for Preparing Students to Understand Complexity*. Dissertation, Interdisciplinary Ecology, University of Florida. Retrieved October 18, 2012, from http://ufdcimages.uflib.ufl.edu/UF/E0/01/71/06/00001/plate_r.pdf.

Plate, R. (2010). Assessing individuals' understanding of nonlinear causal structures in complex systems. *System Dynamics Review*, 26(1):19–33. Available from: http://onlinelibrary.wiley.com/doi/10.1002/sdr.432/abstract.

Podleschny, N. (2008). Playing urban sustainability: the ecology of a simulation game. In *Proceedings of the 20th Australasian Conference on Computer-Human Interaction: Designing for Habitus and Habitat*, OZCHI '08, page 231–234, New York, NY, USA. ACM. Available from: http://doi.acm.org/10.1145/1517744.1517749.

Potts, W. K. (1984). The chorus-line hypothesis of manoeuvre coordination in avian flocks. *Nature, Published online: 24 May 1984*, 309(5966):344–345. Available from: http://dx.doi.org/10.1038/309344a0.

Prince, M. (2004). Does active learning work? a review of the research. *Journal of Engineering Education*, 93(3):223–232. Available from: http://www.jee.org/2004/july/800.pdf.

Quest to Learn (2011). Quest to learn. Retrieved May 01, 2012, from http://q2l.org/.

Quinn, A. J. and Bederson, B. B. (2011). Human computation: a survey and taxonomy of a growing field. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11, page 1403–1412, New York, NY, USA. ACM. Available from: http://doi.acm.org/10.1145/1978942.1979148.

Raia, F. (2005). Students' understanding of complex dynamic systems. *Journal of Geoscience Education*, 53(3):297.

Rettberg, S. (2008). Corporate ideology in world of warcraft. In *Digital culture, play, and identity: a World of Warcraft reader*. MIT Press, Cambridge, MA.

Reynolds, C. W. (1987). Flocks, herds and schools: A distributed behavioral model. *ACM SIGGRAPH Computer Graphics*, 21(4):25–34. Available from: http://portal.acm.org/citation.cfm?doid=37402.37406.

Richmond, B. (2001). *An introduction to systems thinking*. ISEE Systems.

Ritterfeld, U., Cody, M. J., and Vorderer, P. (2009). *Serious Games: Mechanisms and Effects*. Taylor & Francis.

Ross, J. (2011). Pervasive negabehavior games for environmental sustainability. In *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems*, CHI EA '11, page 1085–1088, Atlanta, Georgia, USA. ACM ID: 1979694. Available from: http://doi.acm.org/10.1145/1979602.1979694.

Ross, J., Amsel, N., Beckman, R., and Tomlinson, B. (2010a). EcoPath: adding spatial, social, and gaming contexts to personal tracking systems. Retrieved October 18, 2012, from http://www.ics.uci.edu/~jwross/pubs/SocialCode-2010-01.pdf.

Ross, J., Irani, L., Silberman, M. S., Zaldivar, A., and Tomlinson, B. (2010b). Who are the crowdworkers?: shifting demographics in mechanical turk. In *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems*, pages 2863–2872, Atlanta, Georgia, USA. ACM. Available from: http://dx.doi.org/10.1145/1753846.1753873.

Ross, J., Shantharam, N., and Tomlinson, B. (2010c). Collaborative filtering and carbon footprint calculation. In *Sustainable Systems and Technology (ISSST), 2010 IEEE International Symposium on*, pages 1–6. Available from: http://dx.doi.org/10.1109/ISSST.2010.5507742.

Ross, J. and Tomlinson, B. (2010). How games can redirect humanity's cognitive surplus for social good. *Computers in Entertainment*, 8(4):1–4. Available from: http://dx.doi.org/10.1145/1921141.1921145.

Ross, J. and Tomlinson, B. (2011). Negabehaviors and environmental sustainability. *Journal of Sustainability Education*, 2.

Ryan, R. M. and Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, 25(1):54–67. Available from: http://dx.doi.org/10.1006/ceps.1999.1020.

Salen, K. (2007). Gaming literacies: A game design study in action. *Journal of Educational Multimedia and Hypermedia*, 16(3):301.

Salen, K. and Zimmerman, E. (2003). *Rules of play: Game design fundamentals*. MIT Press.

Saunders, H. D. (1992). The khazzoom-brookes postulate and neoclassical growth. *The Energy Journal*, 13(4):130–148.

Sawyer, B. and Smith, P. (2008). Serious games taxonomy. In *Presentation at The Serious Games Summit at the Game Developers Conference 2008*.

Scavarda, A. J., Bouzdine-Chameeva, T., Goldstein, S. M., Hays, J. M., and Hill, A. V. (2006). A methodology for constructing collective causal maps. *Decision Sciences*, 37(2):263–283. Available from: `http://onlinelibrary.wiley.com/doi/10.1111/j.1540-5915.2006.00124.x/full`.

Schaeffer, S. E. (2007). Graph clustering. *Computer Science Review*, 1(1):27–64. Available from: `http://dx.doi.org/10.1016/j.cosrev.2007.05.001`.

Schaffer, J. (2008). The metaphysics of causation. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Fall 2008 edition. Available from: `http://plato.stanford.edu/archives/fall2008/entries/causation-metaphysics/`.

Schell, J. (2008). *The Art of Game Design: A book of lenses*. Morgan Kaufmann.

Senge, P. M. (1992). *The Fifth Discipline: The Art and Practice of the Learning Organization*. Doubleday, New York.

Senge, P. M. and Sterman, J. D. (1992). Systems thinking and organizational learning: Acting locally and thinking globally in the organization of the future. *European Journal of Operational Research*, 59(1):137–150. Available from: `http://www.sciencedirect.com/science/article/pii/037722179290011W`.

Shaffer, D. W. (2006). Epistemic frames for epistemic games. *Computers & Education*, 46(3):223–234. Available from: `http://dx.doi.org/10.1016/j.compedu.2005.11.003`.

Shaffer, D. W., Squire, K. R., Halverson, R., and Gee, J. P. (2005). Video games and the future of learning. *Phi delta kappan*, 87(2):104.

Sheldon, L. (2011). *The Multiplayer Classroom: Designing Coursework as a Game*. Course Technology Press, Boston, MA, United States, 1st edition.

Shermis, M. D. and Burstein, J. C. (2003). *Automated Essay Scoring: A Cross Disciplinary Perspective*. Psychology Press.

Shiraishi, M., Washio, Y., Takayama, C., Lehdonvirta, V., Kimura, H., and Nakajima, T. (2009). Using individual, social and economic persuasion techniques to reduce CO2 emissions in a family setting. In *Proceedings of the 4th International Conference on Persuasive Technology - Persuasive '09*, page 1, Claremont, California. Available from: `http://dx.doi.org/10.1145/1541948.1541967`.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2):4–14. ArticleType: research-article / Full publication date: Feb., 1986 / Copyright © 1986 American Educational Research Association. Available from: http://dx.doi.org/10.2307/1175860.

Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. In *Computer games and instruction*, page 503–524. Information Age Publishers, Charlotte, NC. Available from: http://myweb.fsu.edu/vshute/pdf/shute%20pres_h.pdf.

Shute, V. J., Masduki, I., and Donmez, O. (2010). Conceptual framework for modeling, assessing and supporting competencies within game environments. *Technology, Instruction, Cognition, and Learning*, 8(2):137–161. Available from: http://myweb.fsu.edu/vshute/pdf/TICL2010.pdf.

Shute, V. J., Ventura, M., Bauer, M., and Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning. *Serious games: Mechanisms and effects*, page 295–321.

Shute, V. J., Ventura, M., and Kim, Y. J. (2011). *Synthesis report on the games, learning, and assessment (GLA) Workshop*. Gates and MacArthur Foundations. Available from: http://myweb.fsu.edu/vshute/pdf/GLA%20report.pdf.

Silberman, M. S., Ross, J., Irani, L., and Tomlinson, B. (2010). Sellers' problems in human computation markets. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, HCOMP '10, page 18–21, New York, NY, USA. ACM. Available from: http://doi.acm.org/10.1145/1837885.1837891.

Sliney, A. and Murphy, D. (2011). Using serious games for assessment. In *Serious Games and Edutainment Applications*, pages 225–243. Springer London.

Sniderman, S. (1999). Unwritten rules. *The Life of Games*.

Sowa, J. F. (2000). Processes and causality. In *Extract from: Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Brooks/Cole, Pacific Grove, CA.

Squire, K. (2002). Cultural framing of computer/video games. *Game studies*, 2(1):90.

Squire, K. (2003). Video games in education. *International Journal of Intelligent Simulations and Gaming*, 2:49–62.

Squire, K. and Jenkins, H. (2004). Harnessing the power of games in education. *Insight*, 3(1):5–33. Available from: http://website.education.wisc.edu/kdsquire/manuscripts/insight.pdf.

Sterman, J. D. (1989). Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment. *Management Science*, 35(3):321–339.

Sterman, J. D. (1994). Learning in and about complex systems. *System Dynamics Review*, 10(2-3):291–330. Available from: http://onlinelibrary.wiley.com/doi/10.1002/sdr.4260100214/abstract.

Sterman, J. D. and Booth Sweeney, L. (2002). Cloudy skies: assessing public understanding of global warming. *System Dynamics Review*, 18(2):207–240. Available from: http://onlinelibrary.wiley.com/doi/10.1002/sdr.242/abstract.

Stokes, B., Duncombe, S., and Fullerton, T. (2010). Direct action games: Games meet the real world. Presented at the Panel at Games for Change Festival 2010, New York, NY. Retrieved October 18, 2012, from http://vimeo.com/13582703.

Strong, A. L., Cullen, J. J., and Chisholm, S. W. (2009). Ocean fertilization: science, policy and commerce. *Oceanography*, 22(3):236–61.

Suits, B. H. (1978). *The Grasshopper: Games, Life, and Utopia*. University of Toronto Press.

Surowiecki, J. (2005). *The Wisdom of Crowds*. Anchor.

Sweetser, P. and Wyeth, P. (2005). GameFlow: a model for evaluating player enjoyment in games. *Comput. Entertain.*, 3(3):3–3. Available from: http://doi.acm.org/10.1145/1077246.1077253.

Terry, L., Roitch, V., Tufail, S., Singh, K., Taraq, O., Luk, W., and Jamieson, P. (2009). Harnessing human computation cycles for the FPGA placement problem. In *ERSA 09*, pages 188–194. CSREA.

Thelwall, M. (2000). Computer-based assessment: a versatile educational tool. *Computers & Education*, 34(1):37–49. Available from: http://dx.doi.org/10.1016/S0360-1315(99)00037-8.

Thomas, C. D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., Erasmus, B. F. N., Siqueira, M. F. d., Grainger, A., Hannah, L., Hughes, L., Huntley, B., Jaarsveld, A. S. v., Midgley, G. F., Miles, L., Ortega-Huerta, M. A., Peterson, A. T., Phillips, O. L., and Williams, S. E. (2004). Extinction risk from climate change. *Nature*, 427(6970):145–148. Available from: http://dx.doi.org/10.1038/nature02121.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological review*, 55(4):189.

Tomlinson, B. (2010). *Greening through IT: information technology for environmental sustainability*. MIT Press.

Tomlinson, B. and Black, R. (2011a). The causality project. Grant proposal submitted to NSF SoCS 2011.

Tomlinson, B. and Black, R. (2011b). Energy education through the causality project. Constellation Energy Funded Grant.

Tomlinson, B., Silberman, M. S., Patterson, D., Pan, Y., and Blevis, E. (2012). Collapse informatics: augmenting the sustainability & ict4d discourse in hci. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, pages 655–664. ACM. Available from: http://dx.doi.org/10.1145/2207676.2207770.

Tomlinson, B., Silberman, M. S., and White, J. (2011). Can more efficient IT be worse for the environment? *Computer*, 44(1):87–89. Available from: `http://dx.doi.org/10.1109/MC.2011.10`.

Tomlinson, B., Yau, M. L., Baumer, E., Ross, J., Correa, A., and Ji, G. (2009). Richly connected systems and multi-device worlds. *Presence: Teleoperators & Virtual Environments*, 18(1):54–71. Available from: `http://dx.doi.org/10.1162/pres.18.1.54`.

Torres, R. J. (2009). *Learning on a 21st Century Platform: Gamestar Mechanic as a Means to Game Design and Systems-thinking Skills Within a Nodal Ecology*. Dissertation, New York University.

Travis, R. (2009). Operation KTHMA (the roleplaying course). Retrieved October 18, 2012, from `http://livingepic.blogspot.com/2009/09/operation-kthma-post-hub.html`.

Tuite, K., Snavely, N., Hsiao, D.-Y., Smith, A. M., and Popović, Z. (2010). Reconstructing the world in 3D: bringing games with a purpose outdoors. In *Proceedings of the Fifth International Conference on the Foundations of Digital Games*, FDG '10, page 232–239, New York, NY, USA. ACM. Available from: `http://doi.acm.org/10.1145/1822348.1822379`.

Turkle, S. (1984). *The Second Self: Computers and the Human Spirit*. Simon & Schuster, New York.

von Ahn, L. (2006). Games with a purpose. *Computer*, 39(6):92 –94.

von Ahn, L. and Dabbish, L. (2004). Labeling images with a computer game. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '04, page 319–326, New York, NY, USA. ACM. Available from: `http://doi.acm.org/10.1145/985692.985733`.

von Ahn, L. and Dabbish, L. (2008). Designing games with a purpose. *Commun. ACM*, 51(8):58–67. Available from: `http://dx.doi.org/10.1145/1378704.1378719`.

von Bertalanffy, L. (1969). *General systems theory: Foundations, Development, Applications*. New York: George Braziller, Inc.

von Bertalanffy, L. (1972). The history and status of general systems theory. *The Academy of Management Journal*, 15(4):407–426.

Waldrop, M. M. (1993). *Complexity: The Emerging Science at the Edge of Order and Chaos*. Simon & Schuster.

Wiener, N. (1948). *Cybernetics or control and communication in the animal and the machine*. John Wiley & Sons, Inc., New York.

Wilson, D. (2011). Brutally unfair tactics totally OK now: On self-effacing games and unachievements. *Game Studies*, 11(1).

Wilson, E. O. (1999). *The diversity of life*. W. W. Norton & Company.

Windeknecht, T. G. (1967). Mathematical systems theory: Causality. *Theory of Computing Systems*, 1(4):279–288. Available from: http://www.springerlink.com/content/k34088w720r23325/abstract/.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press.

Yao, M. (2008). *Visualizing causality in context using animation*. Master's thesis, Simon Fraser University.

Yee, N. (2006a). The demographics, motivations, and derived experiences of users of massively multi-user online graphical environments. *Presence: Teleoperators and Virtual Environments*, 15(3):309–329. Available from: http://dx.doi.org/10.1162/pres.15.3.309.

Yee, N. (2006b). The labor of fun: How video games blur the boundaries of work and play. *Games and Culture*, 1(1):68–71. Available from: http://dx.doi.org/10.1177/1555412005281819.

Zagal, J. P., Rick, J., and Hsi, I. (2006). Collaborative games: Lessons learned from board games. *Simulation & Gaming*, 37(1):24 –40. Available from: http://dx.doi.org/10.1177/1046878105282279.

Zapata-Rivera, J. D., Neufeld, E., and Greer, J. E. (1999). Visualization of bayesian belief networks. In *Proceedings of IEEE Visualization'99, Late Breaking Hot Topics*, page 85–88.

Games Cited

America's Army series. U.S. Army. 2002 - 2012

Call of Duty: Modern Warfare series. Infinity Ward. 2007 - 2011.

Chess (board game). Traditional. 6th century.

Chore-Wars. Kevan Davis. 2007

ElectroCity. Genesis Energy. 2007

EnergyVille. The Economist Group. 2011.

EVOKE. World Bank Institute. 2010.

Foldit. University of Washington. 2008.

Lemmings. Psygnosis. 1991.

Mouse Trap (board game). Ideal. 1963.

Portal. Valve Corporation. 2007.

Re-Mission. HopeLab. 2006

SimCity series. Maxis. 1989 - 2003.

StarCraft series. Blizzard Entertainment. 1998 - 2010

Super Smash Bros.. Nintendo. 1999.

Tom Clancy's Splinter Cell. Ubisoft Montreal. 2002.

Wiki Golf. creator unknown. 2007.

World of Warcraft. Blizzard Entertainment. 2004.