University of Vermont

ScholarWorks @ UVM

2020

Complex systems analysis in selected domains: animal biosecurity & genetic expression

Luke Trinity
University of Vermont

Recommended Citation

COMPLEX SYSTEMS ANALYSIS IN SELECTED DOMAINS: ANIMAL BIOSECURITY & GENETIC EXPRESSION

A Thesis Presented

by

Luke Thomas Trinity

to

The Faculty of the Graduate College

of

The University of Vermont

In Partial Fulfillment of the Requirements
for the Degree of Master of Science
Specializing in Complex Systems & Data Science

January, 2020

Defense Date: November 15th, 2019
Thesis Examination Committee:

Scott C. Merrill, Ph.D., Advisor
William F. Louisos, Ph.D., Chairperson
Hosna Jabbari, Ph.D.
Christopher Koliba, Ph.D.
Cynthia J. Forehand, Ph.D., Dean of Graduate College

ABSTRACT

I first broadly define the study of complex systems, identifying language to describe and characterize mechanisms of such systems which is applicable across disciplines. An overview of methods is provided, including the description of a software development methodology which defines how a combination of computer science, statistics, and mathematics are applied to specified domains. This work describes strategies to facilitate timely completion of robust and adaptable projects which vary in complexity and scope. A biosecurity informatics pipeline is outlined, which is an abstraction useful in organizing the analysis of biological data from cells. This is followed by specific applications of complex systems study to the fields of animal biosecurity and genetic expression. I provide evidence that social cues need to be considered by livestock facility managers in order to increase disease-resiliency of agricultural systems. I also identify significant changes in genetic expression from recent experiments which are advancing the frontiers of regenerative medicine. Areas of future work are discussed including issues related to agriculture and water quality, as well as studies of human behavior and risk perception using experimental gaming simulations.

"Mountains should be climbed with as little effort as possible and without desire. The reality of your own nature should determine the speed. If you become restless, speed up. If you become winded, slow down. You climb the mountain in an equilibrium between restlessness and exhaustion. Then, when you're no longer thinking ahead, each footstep isn't just a means to an end but a unique event in itself. This leaf has jagged edges. This rock looks loose. From this place the snow is less visible, even though closer. These are things you should notice anyway. To live only for some future goal is shallow. It's the sides of the mountain which sustain life, not the top. Here's where things grow."

- Robert M. Pirsig
Zen and the Art of Motorcycle Maintenance: An Inquiry Into Values

ACKNOWLEDGEMENTS

I am so grateful to have had the opportunity to complete a thesis, and it would not have been possible without contributions from countless individuals. To everyone at the SEGS lab, it has been such an amazing experience working with you over the past few years. The intelligence and creativity that the team brings to the table is truly inspiring, and I look forward to continuing collaboration on interesting and valuable research in the future. To my advisor, Scott Merrill, who some have called the best boss ever; thank you for everything you have taught me, and also for giving me the space to grow and develop skills that empower me in tackling any task in front of me. To our other two fearless lab co-directors, Asim Zia and Chris Koliba, your knowledge and perspective have been invaluable in guiding my projects, helping me to frame and position my work in a way that makes it much more relevant and useful. As a committee member, Chris' guidance has had an immeasurable impact on the quality of this work, for which I am extremely grateful. To our data guru, Eric Clark, you tackle extremely difficult problems with consistency and efficiency that is exceedingly rare; thank you for your help on so many different projects, past present and future. To Gabriela Bucini, your positivity and outlook always lifts my spirits, regardless of the situation or lack of natural light in a room. To our animal biosecurity project director, Julie Smith, thank you for supporting me in my pursuit of graduate studies at the intersection of very different topic areas, without you this would not have been possible.

To my committee member Hosna Jabbari, thank you for supporting me as I dive into a new subject area of bioinformatics; I look forward to continuing to learn from you and contribute to exciting new research. To my chair, William Louisos, it has

been so great to reconnect with you through this process, thank you for contributing your perspective to this work. To my family, especially my parents, thank you for always being there for me, I wouldn't be the man I am today with you. Finally, to my friends, thank you for keeping me sane regardless of what I was going through.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

Science is the systematic study of the structure and behavior of the world through observation and experiment. Through the course of our evolution as a species we have uncovered underlying principles which are essential to the practice of science, often characterized as 'laws of nature'. Many of these laws reveal some of the simplest truths about the universe. However, it is believed that there are equally simple principles that comprehensibly describe some of the most complex systems we have discovered. The emerging field of study that is complex systems seeks to identify and extract the mechanisms of these systems; to uncover patterns and correlations that provide explanation and insight in an intelligible way.

Complex systems are generally identified as having some objective or function, either by design or endogenously, originating from within the system itself. The purpose of studying these systems is to extract general principles, commonly delineated as relationships between properties. It can be an extremely difficult task to determine the possible causes of a relationship within a complex system, because of the sheer amount of interacting parts. Recurring themes including spontaneous order, feedback loops or emergence can be identified in the study of complex systems across disciplines and in problems of varying scope. Addressing complexity in a more general sense provides the opportunity to adapt tools built for a specific domain to be universally applicable, a concept known as cross-disciplinary fertilization. [1]

Complex systems are defined as having highly variable behavior and containing many mutually interacting parts. [2] To study these systems, each part can be described individually, as well as in relation to one another. It is the complex nature of these relationships that necessitates researchers in this field to utilize a combination of statistics, mathematics, and computer science; with an emphasis placed on a subset

of computer science known as data science: methods of data collection, manipulation, and visualization. Harnessing computational power and large sources of information allows for analyzing and understanding complex systems in ways that would otherwise be impossible. Within this work, cutting-edge methods are implemented including algorithm design and scripting geared towards biological data, user interface and simulation design related to animal biosecurity questions, and non-trivial statistical modeling applications. The novel conclusions presented here are the product of applying complex systems research methods to existing datasets or application areas, which from a methodological standpoint serves to advance how these tools are utilized in the future.

The methods employed in this work are unique because of their interdisciplinary nature, requiring knowledge from a multitude of subject areas. For example, a knowledge base of psychology as well as animal health was fundamental in the methods employed in Chapter 3. This represents a uniquely complex systems approach: rapidly acquiring knowledge of a domain area in order to identify which tools to utilize. In this case, the problem was structured in an open-ended way such that without the use of complex systems methods, a usable output or solution would have taken much longer or never been achieved. Connecting the uniqueness of the methods to research in the domain of genetic expression, here again the problem structure necessitated a careful choice of how to apply mathematics and scripting. In this instance, the collaboration with biology domain experts could only be advanced by rapidly acquiring domain knowledge and applying statistics and other tools in a proper way.

As a complex systems researcher, applying novel tools and methods to an existing dataset or problem often necessitates an extended literature review. Reading

and becoming immersed in a literature base is vital in order to gain familiarity and comprehension of different systems. In the process of identifying the many parts or elements within a complex system, a natural question to ask is: are the parts themselves complex? If simple parts interact in a system that is characterized as complex on an aggregate scale, or vice versa, it is referred to as emergent behavior. The key idea behind the concept of emergence is that the behavior of the system cannot simply be inferred from the behavior of its components. To expand upon the concept of emergence, it is important to distinguish between local and global emergence. Local emergence refers to a more limited class of emergent behaviors, ones that appear in a small part of the system. Global emergence refers to collective behavior of the system as a whole; and cannot be studied by looking only at individual parts without context, a process known as reductionism. In the study of true complex systems, isolating a part of the system will not provide any information about interaction with other parts. It is therefore the job of the complex systems researcher to gain a comprehensive understanding of a system, in order to study it at an aggregate scale and understand the causes of globally emerging properties.

A logical question to ask when delineating a specific system is: how complex is it? Specifically, how much information is needed to describe the behavior of the system, or what level of detail and granularity is required? A good indicator that the system is indeed complex is if part of the system is removed and the behavior of the entire system changes. The interdependent nature of parts can allow a complex system to be studied; by removing a piece of the system and identifying the resulting change in global behavior. [1] At the boundary of any system is a two-way interaction, with the system influencing its environment and vice-versa. Within a system, elements occupy

space and form structures. There is a temporal aspect to the dynamic processes of interaction between elements in a system, as well as in response to the environment. Discussions of system organization are rooted in the definable purpose or objective that characterizes a system, and the overall function of the system is often intrinsically related to the dynamic or developmentally guided processes that created it.

After defining in general terms what a complex system is, the discussion moves to how these systems are studied. Within the next chapter, a software development methodology will be defined that can be used to study complex systems. In addition, abstractions will be introduced that are helpful in complex systems research in specific domains. Following the overview of methods are two chapters containing applications of the method in specific domains: animal biosecurity and genetic expression. The goal of framing the applications in this way is to identify the patterns and themes that are consistent in the different domains, in order to unify and contextualize the higher-level process that underlies the complex systems research.

CHAPTER 2

OVERVIEW OF METHODS

Embracing a focus on design strategies is critical to successfully achieving project goals. Principal design decisions, known as architecture, permeate all aspects of a software system. These decisions include high-level building blocks like components and configuration, as well as deployment strategies, non-functional properties, and evolution patterns. Non-functional properties encapsulate how a device or interface will be built and maintained, as opposed to what the device does. Placing an emphasis on architecture and design of a system is often correlated with higher quality software, and more effective, low cost solutions. [3]

In addition to the efficiency and reusability that comes with embracing a focus on design strategies, a more specific abstraction is useful in discussion of methods: design patterns. A design pattern encompasses a reoccurring problem, the solution to the problem, and the consequences or trade-offs of application. The goal of enumerating each aspect of a design pattern is to construct a useful common structure that can be applied in many different scenarios. Depending on expertise, one individual's design pattern could be extremely basic to another individual, context dictates the level of

6

granularity. The choice of platform and programming language are extremely relevant in defining the context of a design pattern. [4]

While there will be many design patterns within complex systems research, one of the most important is the general process of developing software. For the purpose of defining the software development methodology within this work, many principles which draw from the Agile methodology [5] have been included. Agile represents an iterative process of development where individuals are emphasized over tools. Important aspects of this framework include being open to changing requirements throughout the development process, delivering working prototypes frequently, and communicating effectively with collaborators across disciplines. Prototypes or preliminary results are project outputs that are not final.

Another literature base which provides an excellent foundation for defining the methods of studying complex systems is the field of human-computer interaction. In addition to the frequent interplay between developers and computational resources in the study of complex systems, many intermediate and final outputs of research can also be guided by principles derived from the study of user interfaces. [6] A software development methodology to describe action steps in complex systems research and the sequence in which they are taken is visualized here using a flow diagram (Figure 2.1).
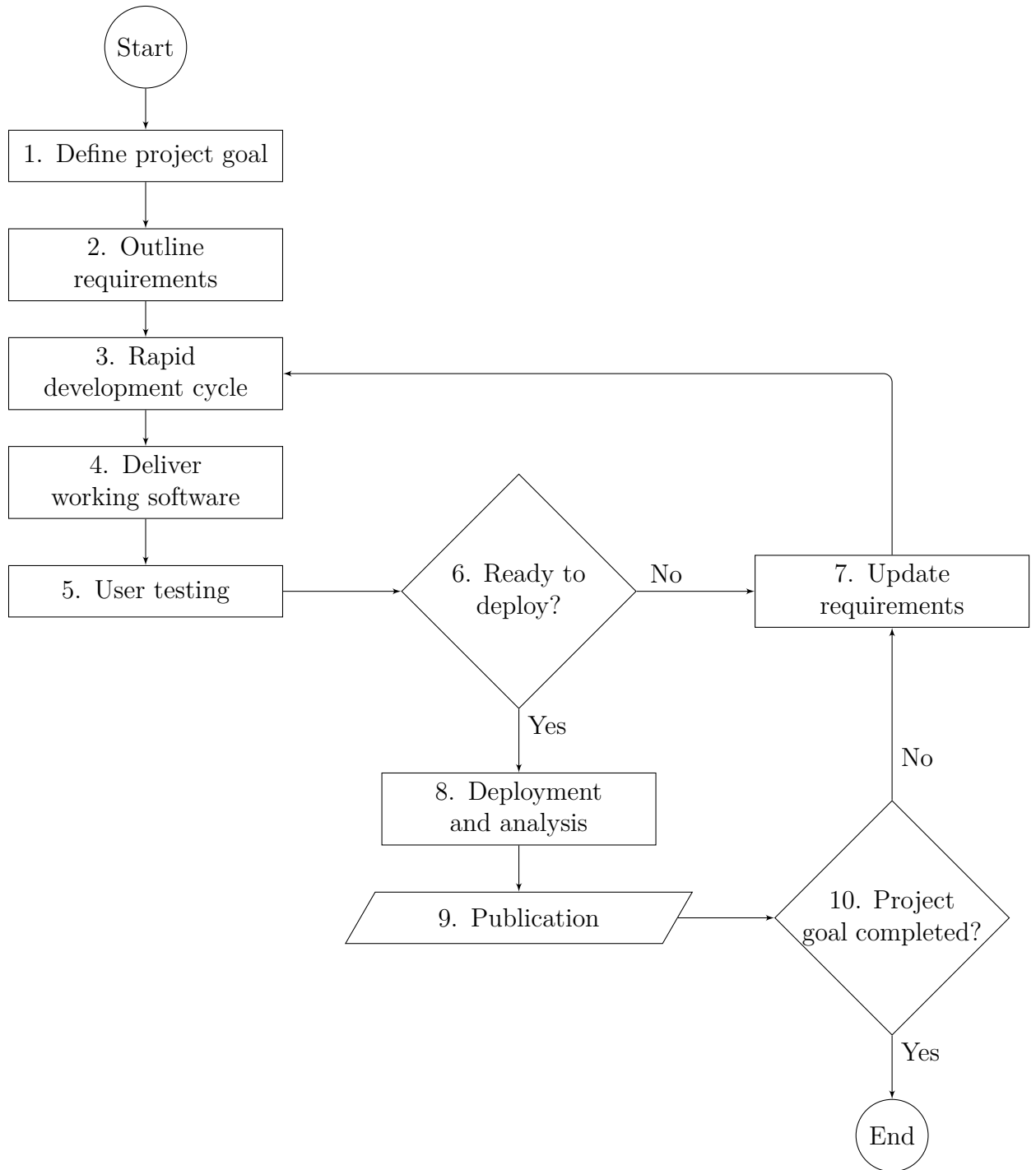
Figure 2.1: Software Development Methodology. Circles delineate initial and terminal nodes, rectangles dictate actions, diamonds correspond with decisions, and parallelograms enumerate an output.

In the first step of the software development methodology, outlining a project goal, there is a need for both creativity and pragmatism to generate testable hypothesis. Working within a specific problem area, focus groups and discussion between individuals from different backgrounds can be extremely beneficial to generate the correct questions that will advance the literature base. It can also be helpful to utilize divergent and convergent brainstorming, which gives space to allow for novel ideas to grow independently before they are weeded out prematurely. In addition to defining project goals, it is the role of the developer or designer to establish a thorough understanding of who the users of a device or software will be.

After a goal is set, the second step in the methodology is to outline requirements. At this stage it is important to avoid any expectations or preset ideas. Requirements should be the bridge between project goals and the desired concrete exchanges between the user and the system. Working backward from desired outputs to outline scenarios from a functional point of view allows for the enumeration of all possible user actions and reactions needed to meet a desired system function. To make requirements achievable, tasks are broken into subtasks, each with a target outcome. [7]

After the initial requirements are outlined, the third step in the process is a rapid development cycle. Rapid in this case means days and weeks, not months and years. [5] Understanding when a rapid development cycle is complete can be complicated and highly dependent on context. It could mean bringing a sketch to a participatory design session, or it could be making a final tweak to a user interface near completion. Either of these cases leads to the fourth step, delivering working software. Documentation and records are important when delivering a prototype, both to maintain connections with collaborators, as well as establish proper version

control.

After working software is delivered, the fifth step is user testing. At this stage users or stakeholders carry out representative tasks on early operational versions and provide feedback. It is expected that there will be novel realizations that are pertinent to the current requirements. Individuals exposed to software for the first time will often struggle in ways that an experienced user or developer could never conceptualize. Testing will unveil cognitive and perceptual limitations of the target audience for a software package. For example, older populations are expected to experience more cognitive load due to having relatively less familiarity with technology. [6] User testing is vital because it can be difficult if not impossible to identify the specific failures of a device without testing it on the intended audience. During user testing, hypothesis can be evaluated for viability; potential differences in populations such as variability in memory, decision making, or risk perception can also be identified with respect to the current prototype. This process allows developers to decrease the amount of time a package takes to build or help it function better when deployed.

The next step in the methodology, step 6, is an evaluation of whether the software will be ready to deploy. Based on the user testing, if the software cannot be deployed, proceed to step 7 and update the requirements. Although it is desirable to deliver working software often, it is also beneficial to iterate many times through steps 3-7 in order to eventually produce the necessary output. If it has been determined that the software is ready to deploy, the project proceeds to step 8, deployment and analysis.

Depending on the users identified by the developer, deployment could signify the delivery of a software output, or it could correspond with operationalizing an interface for online data collection. The significance of this action step in the methodology is

the change from the development of a tool, to the use of that tool. If it was determined that the device or software was in fact ready, then it can be put to its intended use. After deploying a device or software package, analysis naturally follows, a way of gaining an understanding of the results produced by the tool using statistics and mathematics. Analysis culminates in a literature review, articulation of key findings, and participation in a collaborative writing process. The completion of deployment and analysis leads to step 9, publication or dissemination of findings, which is an output of the flow process, indicated by a parallelogram in the diagram.

It is important to distinguish a publication from the culmination of a project, and frame it as an intermediate output. Like the desired iterative process that produces a deployable software package or tool, after publication, step 10 questions if the project goal was completed. At this point an evaluation can be made to determine any gaps in achieving the desired goals of the project. If it is determined that the project goal is not completed to satisfaction, flow returns to step 7, and the requirements are updated. While an end node is included in the flow diagram, multiple publications and subsequent reevaluations of project goals are expected before the end of a project. Having now defined the software development methodology, the remaining two sections in the overview of methods will identify abstractions relevant to specific applications of complex systems science presented in Chapters 3 and 4.

USER INTERFACE & SIMULATION DESIGN

A user interface facilitates interaction between a user and some technology. The goal of an interface is to successfully convey the conceptual structure intended by the de-

signer. This allows a user to create a mental model, a concept derived from cognitive psychology, that allows the user to have some expectation for what the result of an action will be. Objects within an interface can help the user create a mental model through properties that in some way convey function to the user, known as affordances. It can be helpful to delineate the different hierarchies utilized by the designer of such a system. Not only is it beneficial to build a common terminology with which to refer to different aspects of an interface, it is also vital to understand the methods for creating any type of visualization or simulation. The technical aspects of this overview are based around the Unity Development Platform (Unity Technologies), which is a physics-based game engine where objects exist in a virtual space.

Interfaces developed using Unity are composed of scenes which are built to contain different objects by the developer. Scenes are the highest-level abstraction within Unity, each scene holds a nested hierarchy of objects to help organize a project. Oftentimes, scenes are used to group different stages of a simulation. For example, a scene will be used to group objects related to a login page. Upon successful login, a different scene containing objects which compose the main functionality of the simulation will be 'loaded'.

Objects can be attached with a wide variety of components that encapsulate any and all features that may be desired from an interface. One example is the Transform component, which contains an associated position in the coordinate space rendered by the physics engine, as well as rotation and scale of the object. In addition to attaching components to objects, objects can also be nested within each other. Figure 2.2 describes the layout of a simplified interface, but it should be noted that in reality the number of objects and components is much higher.
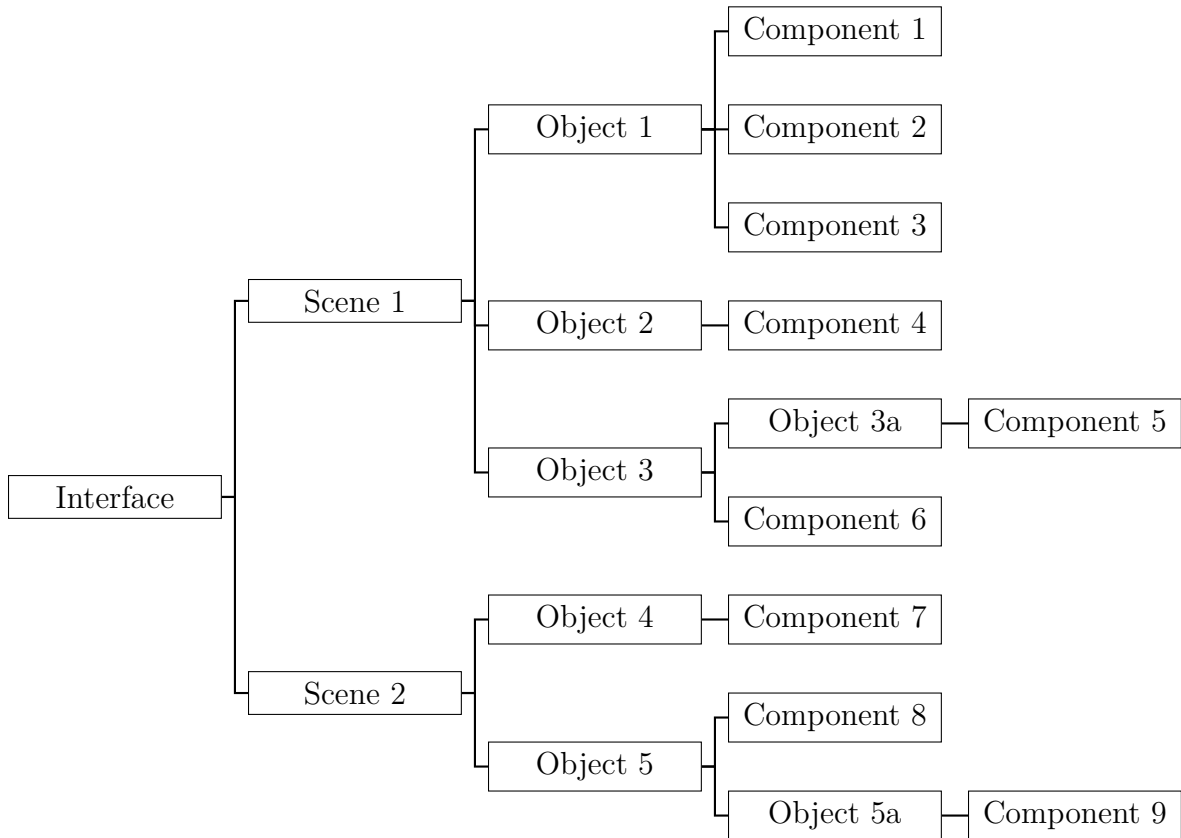
Figure 2.2: Hierarchy of components in a simplified interface with two scenes, seven objects, and nine components.

The Camera component is intuitive to understand, defining a window that is shown to a user when a scene is loaded. The Camera component pairs with another object that has been attached with a Canvas component to deliver a projection to the user interface. The object with the Canvas component is utilized as the parent node for a nested hierarchy of child objects that can each be attached with an Image component. Image components deliver visual information to the user, such as a button or a picture. In the construction of the user interface it is common to utilize objects with two different types of Canvas components in order to generate a layered projection. The first type of Canvas is referred to as 'Screen Overlay', and conveys

information to the user in a menu-style format. The second type of Canvas is referred to as 'Worldview', and operates in a separate coordinate space defined by the physics engine. The Screen Overlay Canvas is structured to be in front of the Worldview Canvas, restricting the users view in a manner controlled by the developer. To explain this concept visually, a user interface is presented in Figure 2.3 as a composition of two types of Canvas components. In this case, the user interface represents a simulated pork production facility, which will be discussed further in Chapter 3.
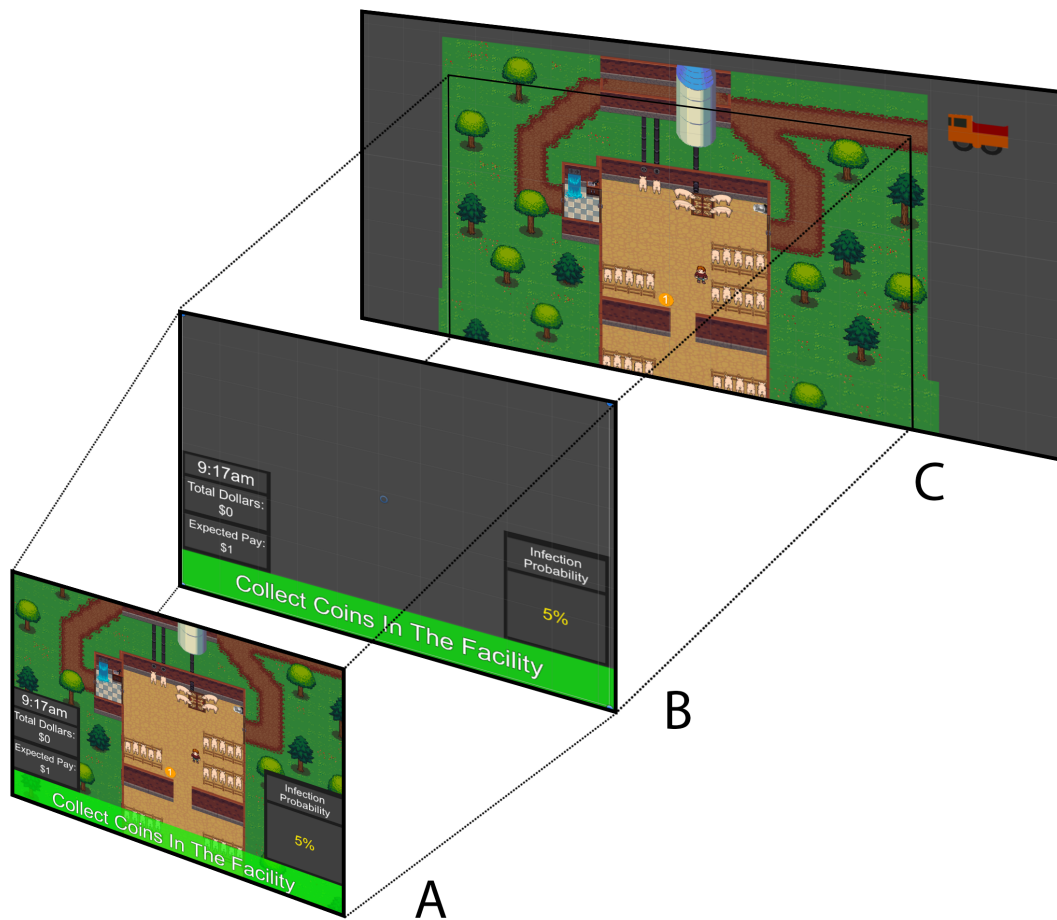
Figure 2.3: Decomposition of a user interface (A) into a Screen Overlay Canvas (B), and a Worldview Canvas (C).

Frame A in Figure 2.3 is representative of exactly what a user of this interface would see. However, from a development standpoint, there are two distinct objects with Canvas components creating this abstraction for the user. The Screen Overlay Canvas (B) does not change in its position or orientation during a simulation. However, the user's view of the Worldview Canvas (C) does change, depending on the location of the Camera component, which can be moved. Another example of this would be in a driving simulation where the environment changes as you drive in the virtual world, but you are given a constant view of your virtual gas gauge. This is analogous to the purpose of the Screen Overlay Canvas, and continuing with the analogy, the virtual world that gives the perception of motion as you drive in a simulation is represented here by the Worldview Canvas (C). Note how the black lines delineate the mapping from the user interface to the two canvas, and how some of the game space which can be seen in (C) is not visible in (A). It is often crucial that the entire world created by the physics engine is not displayed to the user all at once, and this is how the vantage point shown in Frame A is created.

In addition to controlling what the user sees, it is also required in certain implementations that a user is given some control within the simulation. This is achieved using Script components to process external inputs like a keystroke or mouse click. Script components, containing code written in a higher-level language like C# [8], can be attached to any object and are compiled when a scene is 'loaded' or 'run'. Script components can also be triggered when the simulation is in progress. As a user provides input into a system, it is critical that they receive output that is consistent with their mental model of the interface. Some of this feedback is achieved using Collider components, which define a boundary within the coordinate space. Collider

components can either act as an impenetrable barrier that restricts movement within a simulation, or be penetrable and initiate a trigger when the boundary is breached.
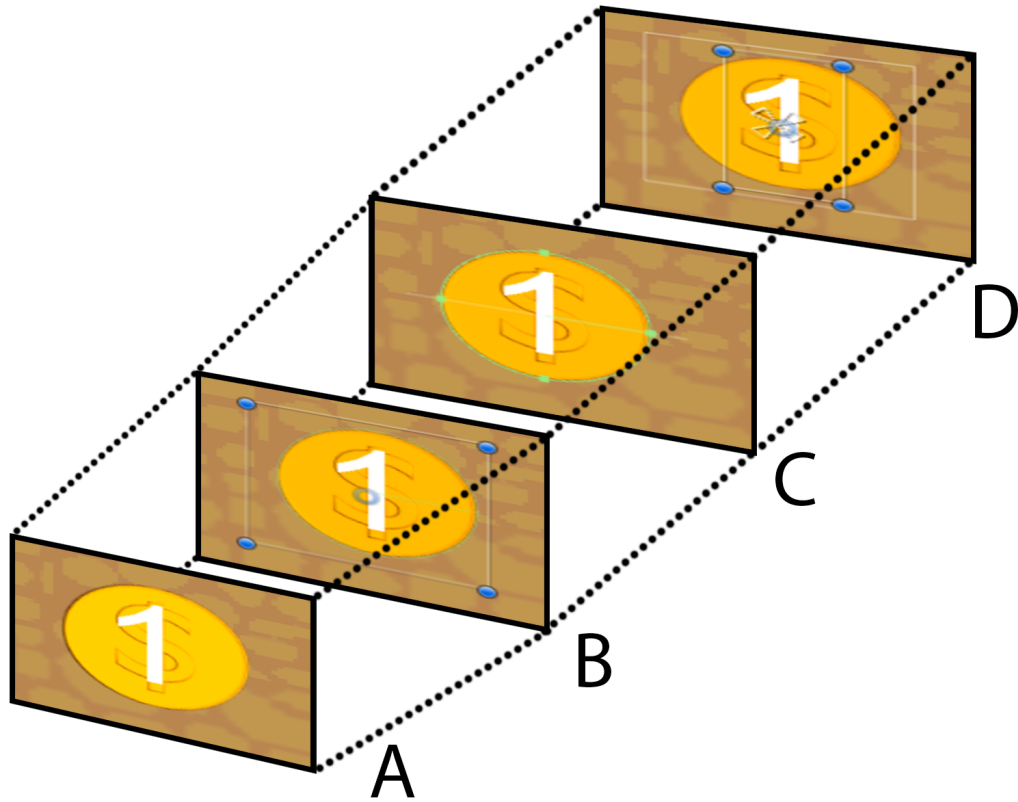
Figure 2.4: Decomposition of a virtual coin (A) into an Image component displaying a coin (B), a Collider component (C), and an Image component displaying a '1' (D). The ordering of the three components (B, C, and D) is not significant.

Figure 2.4 displays the composition of a simulated coin, again the example is drawn from the simulated pork production facility interface that will be discussed in Chapter 3. In Frame A the virtual coin is displayed as it would be shown on the user interface. Frames B and D correspond to the two Image components containing the coin and the '1' text, respectively. Frame C represents the Collider component, shown as a green line. In this implementation the collider was used as a trigger. When

another object breached the boundary signified by the green line, it triggered the coin to be destroyed and another variable representing the user's score to be incremented.

User interfaces and simulations are built using layering and hierarchy to nest scenes, objects, and components into a unified conceptual structure. Interestingly, in the construction of outputs like user interfaces which are highly visual in nature, Scripts are just one type of component that works in conjunction with many others. In other domains of complex systems study that have outputs that do not rely on rendering a space using a physics engine, scripting will play an even more central role.

Bioinformatics Pipeline

Bioinformatics is a field of study devoted to the collection and analysis of biological data, such as genetic codes or sequences. Due to the specific type of processing required for these tasks, analysis in this domain uses a pipeline framework, where files undergo a series of sequential transformations. Pipelines can range in complexity, and often culminate in the generation of analysis summaries and reports. Connecting back to the overall software development methodology, when developing a pipeline it is important to consider the end users. In this context, it is expected that individuals have varying combinations of experience in biological and computer science backgrounds.

In contrast with the hierarchy of different components used to build a user interface, the basic building block of pipelines is restricted to a single component: scripts. It is important to consider the dependencies of a script, which are categorized as upstream or downstream. Upstream dependencies refer to prior scripting outputs

that affect the input to an intermediate script. Downstream dependences classify restrictions on intermediate outputs based on their use as future inputs. Modularity is also extremely important, allowing recovery from checkpoints at different stages in a pipeline of scripts. Aligning closely with modularity is the concept of reentry, which allows the execution of a pipeline from an intermediate step without rerunning previous steps. Interaction between pipelines and resources hosted online are increasingly common. These type of handshakes between a pipeline and an online resource are accomplished using application programming interfaces (APIs), a set of functions that allow access of an application or service remotely. Pipelines that are themselves hosted online enhance reproducibility of an analysis and contribute to future augmentation or collaboration. [9]

A pipeline will now be introduced that is used in an application of complex systems research in Chapter 4. The goal of this pipeline is to identify significant changes in genetic expression. Genetic expression is an essential life process by which information from genes is synthesized into functional proteins. Advancements in technology related to the generation of human pluripotent stem cells, which are cells derived from embryos or fetal tissue, are providing new datasets that can advance the literature base of genetic expression. [10] While induced pluripotent stem cells can be generated from adults, accumulated mutations from aging and UV exposure may limit potential to model diseases or develop novel therapies. [11] Human umbilical vein endothelial cells are preferable for a variety of reasons included their availability through non-invasive methods. [11] Proper analysis of these datasets using a pipeline framework can generate novel results in the characterization of gene expression profiles, with potential applications for modeling and treatment of disease.

The output required from this pipeline is simple when abstracted, given a set of changes in gene expression, determine which of the changes are significant. However, when an input dataset may include tens of thousands of observations, it is crucial to define each sequential step in detail. The first phase in the pipeline is to process raw data and describe in mathematical terms the confidence we have that a significant change in gene expression has occurred. A simple method to do this, referred to as 'avg_raw_change', identifies the mean or average change of expression in a gene, over all observations. Another method of identifying significance is the implementation of a software package 'fcros', which utilizes linear algebra to generate probabilities of significance based on ranked values. This method detects differentially expressed genes based on their fold change rank, which is a ratio of averages from control to test samples. [12] It can also be helpful to examine the range of fold changes on a logarithmic scale outside of the fcros package, this third method of determining significance is referred to as 'log_fold_change'.

After generating higher-level mathematical features from the raw data, the next step in the pipeline is to identify thresholds that define the criteria for significance. It can be helpful at this point to consult with domain experts and use genes that have been determined to be significant in previous research as a guideline. The process of eliminating genes from consideration for analysis is referred to as 'filter_for_significance'.

At this point in the pipeline, it has been determined which of the changes are significant. However, additional steps are required to generate comprehensible analysis reports. Gene Ontology is a structured vocabulary classification system that can be used to sort genes into a hierarchical structure based on their functional

characteristics. [13] Sorting genes based on their functional classification is an integral part of linking the genes together and explaining the key takeaways of the expression changes, referred to here as 'subset_by_ontology'. The final step in the pipeline, 'quadrant_rel_change_sort', is a method of grouping the gene expression data. In this sorting, the paired positive and negative changes of a gene as it is differentiated to either endothelial or neuronal cells are grouped into the traditional mathematical quadrants (1, 2, 3, 4). Completed analysis reports contain genes that have significantly changed in expression, grouped both by their associated functional characteristics as well as the result of the quadrant-based sorting.
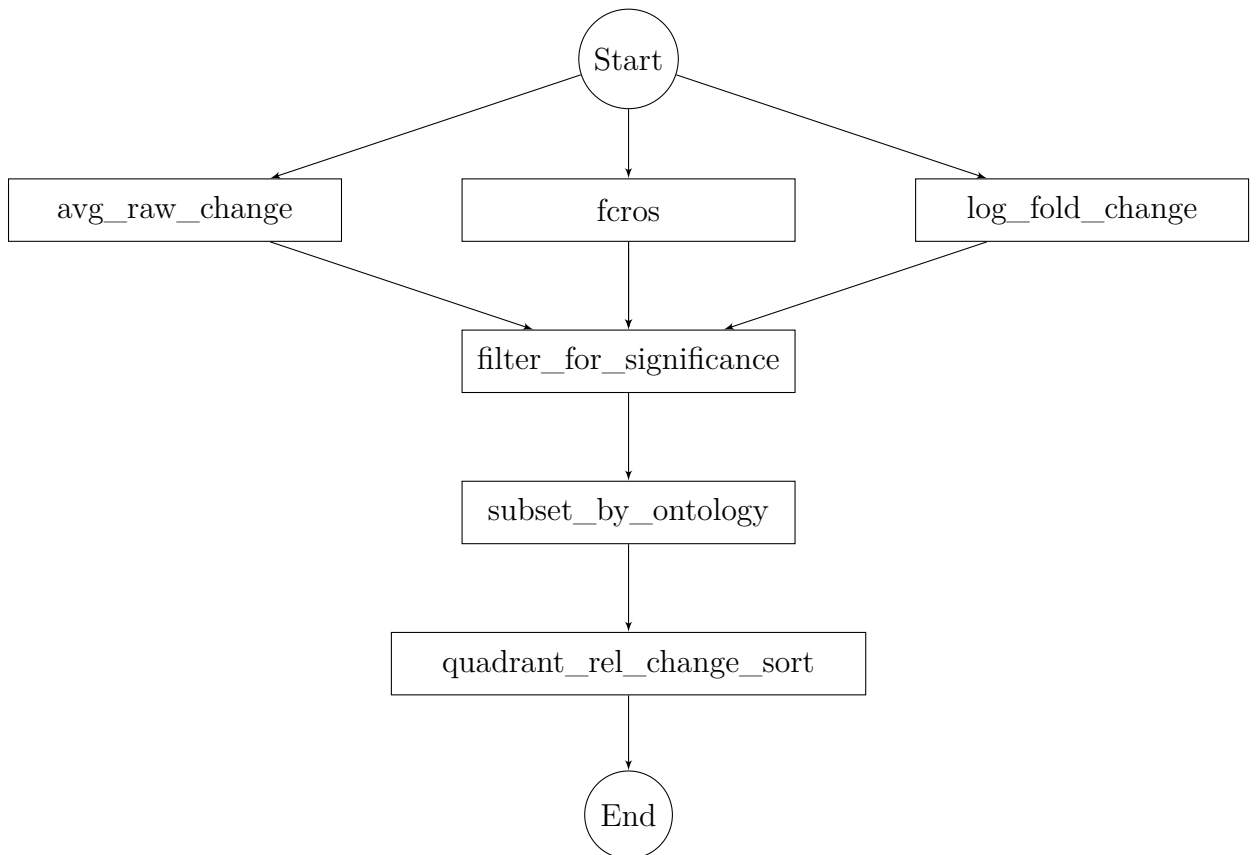
Figure 2.5: Pipeline: Significant Gene Expression

Chapter 3

Effects of Social Cues on Biosecurity Compliance in Livestock Facilities: Evidence from Experimental Simulations

ABSTRACT

Disease outbreaks in U.S. animal livestock industries have economic impacts measured in hundreds of millions of dollars per year. Biosecurity, or procedures intended to protect animals against disease, is known to be effective at reducing infection risk at facilities. Yet to the detriment of animal health, humans do not always follow biosecurity protocols. Human behavioral factors have been shown to influence willingness to follow biosecurity protocols. Here we show how social cues may affect cooperation with a biosecurity practice. Participants were immersed in a simulated swine production facility through a graphical user interface and prompted to make a decision that addressed their willingness to comply with a biosecurity practice. We tested the effect of varying three experimental variables: (1) the risk of acquiring an infection, (2) the delivery method of the infection risk information (numerical versus graphical), and (3) behavior of an automated coworker in the facility. We provide evidence that participants changed their behavior when they observed a simulated worker making a choice to follow or not follow a biosecurity protocol, even though the simulated worker had no economic effect on the participants' payouts. These results advance the understanding of human behavioral effects on biosecurity protocol decisions; demonstrating that social cues need to be considered by livestock facility managers when developing policies to make agricultural systems more disease resilient.

INTRODUCTION

Endemic and emergent diseases remain a constant threat to the animal and economic welfare of the livestock industry. A national survey of U.S. hog producers found that from 2014-2017, 54.3% reported having suffered an outbreak of Porcine reproductive and respiratory syndrome virus (PRRSv), and 43.9% reported undergoing an outbreak of Porcine epidemic diarrhea virus (PEDv). [14] Annual economic losses due to PRRSv and PEDv are estimated at $580.62 million [15], and upwards of $900 million [16], respectively. Animal health issues become even more pronounced considering the current threat of African swine fever, a highly virulent disease that can cause up to 100% fatality in pigs. [17] Furthermore, analysis of consumer preferences indicates that animal welfare is an important attribute to some consumers of livestock products. [18] During disease outbreaks, public concerns related to food safety can cause reductions in pork consumption, which carries economic ramifications for the swine industry. Implementation of biosecurity best management organizational policies is critical to effectively prevent or control outbreaks of existing and emerging virulent diseases. [19]

Human behavioral factors that can influence biosecurity implementation have been identified as crucial to mitigating the risk of outbreaks; although limited knowledge exists on the actual relationship between these factors and behavior. [20, 21] Analysis of human-behavior can be undertaken in contexts of varying scope: strategic, tactical, and operational. [22] Taking a broad view, big picture strategic decisions are guided by long-term objectives. Biosecurity issues at the strategic scale are often related to network interactions of facilities and service providers across production chains. The

emergent behavior of networks, an important consideration of strategic policy makers, is intrinsically related to more localized tactical decisions that are made within a specified region. At the tactical level, farm managers decide whether or not to invest in and implement preventive biosecurity protocols. From a narrower, more localized perspective, operational level biosecurity can be viewed as a continuous series of decisions made by production workers indicating their willingness to follow or comply with biosecurity protocols (hereafter referred to as compliance with biosecurity). [23] The operational, tactical, and strategic levels of biosecurity are interconnected, for example, tactical decision-making influences operational level protocols. And while it is known that operational level compliance with biosecurity will impact the implementation and efficacy of tactical and strategic biosecurity decisions, limited feedback describing operational level behavior is currently available to inform tactical and strategic level decision-making.

Compliance with biosecurity, such as consistently following sanitation protocols before entering a production facility, has been reported to significantly reduce disease. [24–26] Unfortunately, poor compliance with biosecurity is an endemic problem in many animal production systems. A detailed questionnaire of 60 Swedish farrow-to-finish herds found that particular biosecurity measures were applied for visitors in more than half of farms; but those same measures were carried out by farmers and staff themselves on only 32% of farms. [27] Hidden cameras at Quebec poultry farms documented 44 different biosecurity lapses made by workers and visitors over a four-week period. [28] Workers may be exposed to a variety of information about the consequences of a facility infection; but they balance the cost of infection with complacency and a tendency to become lax in day-to-day activities. This relaxation

of biosecurity effort has been examined from a temporal perspective, with evidence suggesting that people view the likelihood and impact of an event to be reduced as time since the event increases. [29–31] Referred to as temporally-based psychological distancing, the farther in the past an event occurs, the less likely and impactful such an event is perceived to be. Pressure to complete work efficiently with time constraints has created scenarios where workers find it unrealistically challenging to complete their job while complying with biosecurity standards. [32] A better understanding of the complex human decision-making process that influences workers' willingness to comply with biosecurity protocols requires innovative research approaches and data collection techniques that can provide novel feedback for tactical and strategic level decision makers.

Studies of human behavioral strategies have applied serious games developed for data collection or education as early as 1962 for designing robot mining simulations. [33] Experimental economic games are a particular type of serious game in which participants are incentivized with monetary payouts. Performance-based incentives are known to increase engagement and salience in decision-making. [34–36] Indeed, Holt & Laury found in their multiple price lottery experiment that risk aversion increased when human subjects were faced with real financial incentives. [37] Computer-based simulations provide a unique opportunity to study the mechanics of decision-making in a controlled environment. [38] Experiments utilizing computer software have had success amongst adolescents, increasing empathy [39], as well as awareness of substance use [40]. Within the domain of animal biosecurity, serious games have explored the effects of information awareness and audience on tactical investments [41, 42], as well as the effects of message delivery method on operational

compliance [43].

Recent research conducted by Merrill et al. [43] utilized an experimental simulation of a livestock production facility to examine factors that may influence perception of disease risk, thus affecting biosecurity compliance. Their experimental treatment variables included: information regarding disease infection risk, uncertainty associated with disease infection risk information, and the message delivery method used to communicate disease infection risk. These factors have been identified as important within the farmer biosecurity decision-making process. [20,44] Research has suggested that farmers in the United States exhibit risk tolerance [45], but they may be more likely to implement or comply with biosecurity as their perceived risk of infection increases [19,41,46]. Indeed, Merrill et al. [43] found that as the actual infection risk within the experimental simulation increased, so did compliance with a biosecurity practice. The certainty of infection risk information is also expected to affect biosecurity implementation and compliance [19]; although the effect may change depending on the domain of interest, i.e. tactical versus operational. Merrill et al. [41] found that in a *tactical* experimental simulation of disease in a swine production region, an increase in disease risk certainty was associated with increased biosecurity implementation. However, in an *operational* experimental simulation of a single livestock production facility, an increase in certainty was associated with decreased biosecurity compliance. [43] The opposite responses to increases in disease risk certainty at the tactical and operational levels highlights the complexity of human behavior, and the need to provide operational feedback to tactical and strategic decision makers which is currently unavailable.

Another factor that is expected to impact compliance with biosecurity protocols is

the message delivery method when communicating disease risk level. [47] The method of delivering a message is important because humans exhibit affective reasoning: feelings and initial reaction to stimuli guide decision-making. [48, 49] Additionally, messaging affects the balance of experiential and analytical reasoning in the processing of statistical information. [50] Humans rely on a limited number of mental heuristics, or ways to reduce problem complexity, when presented with decisions under uncertainty. [51] Visual or graphical communication of risk can be advantageous in attracting and holding attention. [47] Indeed, Merrill et al. [43] found that risk information delivered in a graphical format was more effective in increasing biosecurity compliance, with respect to messages delivered in a numeric or linguistic format.

Contextual and situational factors also determine how probability is understood. [52,53] Generally, humans underestimate risks that occur frequently [54], and discount the risk of an event if the probability is low [55]. However, the exact definition of *low* is malleable and subjective, as it can change depending on context. Humans utilize mental shortcuts when making decisions with limited time. One such common heuristic is referred to as anchoring and adjustment. [51] The following example illustrates how the formulation of a problem can profoundly impact the final answer. When two groups of high school students were asked to estimate a numerical expression within 5 seconds, they used extrapolation and adjustment to formulate their answer under time pressure. The first group, when asked to estimate $1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8$ yielded a median estimate of 512; while the second group, presented with the expression $8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$ yielded a median estimate of 2,250. [51] Although the correct answer is 40,320 in both cases, changing the initial point of reference completely alters the result of the off-hand mental calculation.

In this study we expand upon the work of Merrill et al. [43] by introducing a novel variable as a treatment: information about social cues in a production facility. Response to social cues is grounded in social value orientation, the weight given by an individual to their own welfare, and the welfare of their interaction partner, in evaluation of the desirability of a particular outcome. Predicting behavior is challenging when the same social cue can be interpreted as an incentive or a deterrent. Individuals who assign positive weights to partners' welfare are classified as pro-social. Those who assign negative value or do not assign any value to their partners' welfare are characterized as pro-selves. Where pro-social individuals can interpret a behavior as a reflection of good intentions, pro-self individuals value only their own welfare and may interpret the same behavior from an interaction partner differently, as a sign of weakness. [56] Variability in internal logical and activities that motivate farmers in particular makes a "one-size-fits-all" strategy impossible. [57] Previous work has shown the potential for social cues to contribute to behavioral flexibility, influencing the way individuals react to their environments. Performance based imitation, copying actions of those who are seen to do well, is an observed phenomenon known as social sampling. [58] Social cues within a computer simulated maze experiment were seen to dramatically increase the adoption of novel behavior patterns. [59] Research describing changes in operational level biosecurity compliance behavior as a result of social cues is not currently available. Identifying that social cues have an effect on operational biosecurity compliance could aid decision makers at the tactical and strategic level as they seek to increase compliance with protocols.

We developed a novel experimental variable to treat, based on the framework of the serious game designed by Merrill et al. [43], in which participants were confronted

with a compliance decision related to a common biosecurity practice, usually referred to as showering in-and-out. The shower-in, shower-out biosecurity practice, a component of the line of separation biosecurity protocol, involves changing clothes and showering before entering or exiting areas with livestock to reduce transmission of disease between animals within a facility and the outside environment. The practice is known to be effective at reducing the risk of infection. [24] However, workers may neglect, avoid, or insufficiently complete the practice due to the time it takes to use it. [32]. For example, effectively showering multiple times a day is time consuming and may be perceived to have negative repercussions such as damaging one's hair and skin or inhibiting completion of daily tasks. We developed novel experimental scenarios that tested participants' willingness to comply with the shower biosecurity practice. By using the shower practice, participants incurred an associated time cost, but they avoided the potential risk of an infection. Specifically, we placed individuals into a farming situation where they were asked to make a binary biosecurity decision. Here we hypothesize (H1) that when provided a social cue, in this case a biosecurity behavior of a coworker, individuals may react by mimicking or doing the opposite of the cue; but regardless are more likely to change their behavior in response to the cue, than if they did not receive the cue.

Experimental variables manipulated the following factors: (1) the risk of acquiring an infection, (2) the delivery method of the infection risk, either a numerical (Numeric) value or a graphical threat gauge style image (Graphical), and (3) behavior of an automated coworker in the facility, either demonstrating compliance (Compliance by Coworker), demonstrating non-compliance (Non-Compliance by Coworker), or not demonstrating (Coworker Control). In addition to the three primary drivers of be-

30

havior, infection risk, message delivery method, and social cue; we also looked for an additional secondary driver of behavior, temporally-based psychological distancing. While we expect that participants will increase compliance in response to an infection event, this tendency to comply is likely to decrease as time passes. [29, 31, 55] Referred to as a psychological distance effect, Merrill et al. [43] found that compliance did increase directly after an infection, but the effect decreased with time after the infection. We hypothesized (H2) that compliance would increase directly after an infection event, and temporally-based psychological distancing would occur defined by a decrease in compliance with increasing time since an infection.

Methods

Deployment

We conducted a single experiment to examine human behavioral responses to social cues in a simulated production facility. Participants were recruited using the online workplace Amazon Mechanical Turk, which has been identified as a representative sample for the U.S. population [60], and a viable alternative to traditional data collection [61]. Recruits were informed that their pay would be based on performance during the experiment. Before the experiment commenced, an informational slideshow was displayed explaining the purpose of the study and mechanics of the game. This was followed by a screen allowing the recruit to choose between proceeding to play the game or declining to participate. Institutional Review Board approved practices were followed for an experiment using human participants (University of Vermont IRB #

CHRBSS-16-232-IRB).

Experiment Design

The simulated pork production facility was built using the Unity Development Platform (Unity Technologies, Version 5.6.3) and hosted online using WebGL [62]. Each participant acted as a worker and was provided information in the form of treatments that differed by combination of experimental variables: the risk of infection if they chose not to comply, the delivery method of the infection risk message, and behavior of a coworker present with them in the simulated facility. With the provided treatment information, participants were confronted with a choice to use the shower biosecurity practice or bypass the practice to avoid costs associated with usage of the practice. Each round lasting up to 70 seconds(s) represented one work day, 9am to 6pm. The experiment began with one practice round, followed by 18 rounds of incentivized play.

Figure 3.1: Screenshot of a game round showing infection risk delivered as a numerical (Numeric) message, the participant-controlled worker, automated coworker, coins (internal tasks), the shower biosecurity practice (blue arrow) and emergency exit (red arrow).

To act as a worker in the facility, the participant used the computer keyboard. Each round, the worker began inside the facility. Tasks, represented as spinning coins, appeared every 2 s. When the participant moved their worker to a coin, they earned $1 experimental dollar. Once during each round, a high-value task would appear outside the virtual facility. The value for attending to this outside task was based on the time it took to accomplish, starting at $30 experimental dollars and decreasing by $1 experimental dollar per second. To earn the experimental dollars for completing this high-value task, participants chose to comply and use the shower biosecurity practice which required extra time both exiting and entering the facility (with an approximate observed cost of $8.67 experimental dollars), or avoid compliance by using the emergency exit, which carried no associated time costs but risked infection of the facility's swine.

33

More specifically, across all treatments participants were asked to leave the facility to complete an outside task and were confronted with the decision of how to leave the facility, either by complying with the biosecurity practice or by skipping the practice and leaving through the emergency exit. If participants decided to use the shower biosecurity practice and "comply", participants activated a five-second counter that simulated the time it takes to shower and change clothes. After 5 s the virtual worker could exit the shower and complete the outside task. The same procedure, with another 5 s delay, was repeated upon re-entry of the facility post-task completion. If the participant decided on non-compliance with the shower biosecurity practice, and thus, left the facility through the emergency exit, they incurred no time cost, but there was an associated chance of infection based on the actual infection risk probability during the given round. The risk of using the emergency exit was quantified using the infection information presented to the participant, which varied by treatment (see Table 3.1 for breakdown of information provided to participants by experimental variable and associated treatment levels). If an infection occurred, calculated using a pseudorandom number generator, the round ended immediately, and the participant lost $50 experimental dollars as well as any expected payout they had collected during the round. If an infection did not occur, the round continued until the normal end of the workday. The mean observed opportunity cost of using the shower practice as opposed to the emergency exit was calculated to be $8.67 experimental dollars, due to the time lost that could have been used to complete the outside task more quickly and collect coins inside the facility upon reentry.

In addition to the participant's worker, an automated coworker was included in all simulation scenarios to provide implicit social cues to the player. The only explicit

information given to the participants regarding the coworker was during the pre-game slide show where they read that there may be another worker in the facility with them. The automated coworker was completely predetermined in its actions, exhibiting one of three behaviors in any given round: 1) compliance with the shower biosecurity practice, 2) use of the emergency exit, or 3) not exiting the facility (Control). During the coworker compliance and non-compliance treatments, participants would see the decision made by the coworker, before they made their own decision to use the shower biosecurity practice or the emergency exit. This demonstration by the coworker, an implicit social cue, was intended to be observed by the participant with the potential to affect their decision to comply.

After each round, the number of experimental dollars earned within the round was displayed on the participant's screen. In addition, a cumulative sum of the total experimental dollars earned thus far in the experiment was displayed. At the end of the 18 experimental rounds, participants received $1 U.S. for each $350 experimental dollars plus a base pay of $3.00 U.S.

Experimental Variables and Treatments

Experimental variables were designed to test factors shown to influence human behavior: risk of infection, message delivery method, and social cue. Participants were confronted with infection risk, delivered in different message formats, at the start of each round. This infection risk information was used by the participant during the round when they chose to either comply with the shower biosecurity protocol, or risk using the emergency exit for a slightly higher payout. The infection risk information was delivered numerically (e.g., "5% infection risk"), or graphically using a threat

gauge (Figure 3.2).
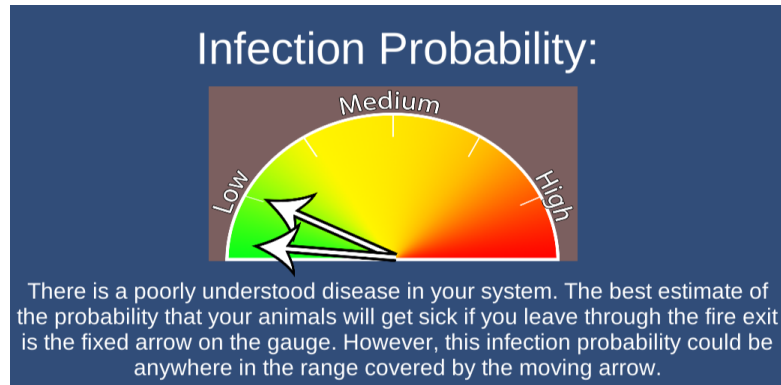
(a)
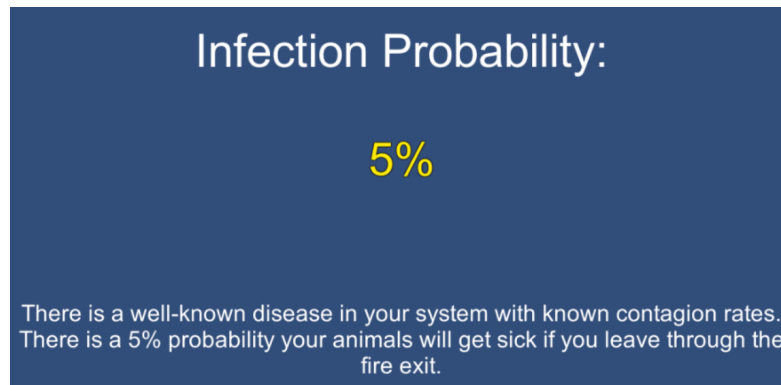
(b)

Figure 3.2: Start of round infection risk delivered as a graphical (Graphical) message; two arrows, one fixed, the other moving, were used to convey a best estimate and uncertainty around that estimate (a). Start of round infection risk delivered as a numerical (Numeric) message (b).

In the original work of Merrill et al. [43], the certainty of the disease risk information and the message delivery method were both treated as separate variables. When disease infection risk was treated with certainty, participants were provided a single infection probability. Disease infection risk was treated with uncertainty by providing a best estimate of the infection probability in addition to a range of potential values. It was determined that numerical values *with certainty* were the most likely to

36

be associated with avoidance of the biosecurity practice, and graphical threat gauge style images *with uncertainty* were the best at increasing willingness to comply. [43] In order to increase our sample size per treatment, we utilized only these two types of message delivery methods: numeric with certainty, and graphical with uncertainty, respectively. Hereafter, these are referred to as Numeric and Graphical.

For our novel social cue experimental variable, an automated coworker demonstrated one of three behaviors prior to the participant's decision to comply. Unbeknownst to the participant, the compliance decision chosen by the coworker was predetermined by treatment. Coworker demonstration of compliance involved the automated coworker using the shower biosecurity practice; likewise, the coworker demonstration of non-compliance involved the coworker using the emergency exit. No demonstration, the control treatment, indicates that the coworker never left the facility.

In summary, the experiment had three infection risk treatments:(very low (1%), to low (5%), to medium (15%)), two infection risk messaging treatments (Numeric and Graphical), and three social cue treatments (Compliance by Coworker, Non-Compliance by Coworker, and Coworker Control). A complete block design was utilized, in which data were collected for every combination of the three experimental variables and their associated levels. Two additional variables were also used in this experiment. First, the psychological distance effect, to identify changes in behavior related to experiencing infection events. The distancing effect was quantified by simply counting the number of rounds since an infection event occurred. Second, because the experiment takes place in a series of rounds, we used a variable referred to as play order to control for within-experiment learning. [63]

Treatment	N
Infection Risk: 1% (Very Low)	648 (6 rounds * 108 participants)
Infection Risk: 5% (Low)	648 (6 rounds * 108 participants)
Infection Risk: 15% (Medium)	648 (6 rounds * 108 participants)
Message Delivery Method: Numeric ("1%," "5%," or "15%")	972 (9 rounds * 108 participants)
Message Delivery Method: Graphical (A threat gauge with arrows used to indicate risk)	972 (9 rounds * 108 participants)
Social Cue: Compliance by Coworker	648 (6 rounds * 108 participants)
Social Cue: Non-Compliance by Coworker	648 (6 rounds * 108 participants)
Social Cue: Coworker Control	648 (6 rounds * 108 participants)

Table 3.1: Experiment treatments.

Analysis

Linear effects

The response variable in our experiment is binary, either the participant used the shower biosecurity practice, or avoided compliance by using the emergency exit. To explain the response variable, a set of mixed-effect logistic regression candidate models were generated using the statistical programming language, R. [64–66] The set of candidate models included mixtures of experimental variables (Table 3.1), interaction terms between experimental variables, as well as predictor variables: (1) psychological distance, and (2) play order. Participant was added as a random effect in all models to account for variation between individuals. To identify the model of best fit, we used an information theoretic approach to test how well each of our candidate models

explains the data. [67,68] Models were evaluated using Akaike's Information Criterion (AIC), where the lowest AIC value indicates the most parsimonious candidate model that best explains variation in the response variable with the fewest parameterized variables. [69]

Ratio of variances

Nonlinear effects resulting from inconsistent responses to the social cue experimental variable are identified in a separate examination of the strategy variance components. Here we quantify strategic variability by calculating how an individual changed their compliance strategy between the three social cue treatments: compliance, non-compliance, and control. Due to our complete block experiment design, we can isolate an individual's average compliance in response to any one of our three social cue treatments by averaging over the other two experimental variables: infection risk and message delivery method. Having identified an individual's average compliance strategy for each of the three social cue treatments they were presented, we then calculate the differences in strategy between each of the three pairs of social cue treatments (compliance vs. non-compliance, compliance vs. control, non-compliance vs. control). For example, how did an individual's average compliance change between all rounds in which they were presented the social cue compliance treatment, versus all rounds they were presented with the social cue non-compliance treatment.

After quantifying individuals' strategic variability using changes in average compliance, we can aggregate all participants' average compliance changes into distributions, grouped by the pairs of social cue treatments between which individuals' may have collectively changed their average behavior. The variance of each of the three

distributions can be evaluated in relation to one-another as ratios of variances, which quantify differences in strategic variability observed over all individuals. Confidence intervals were calculated using an F-test to determine if any of the ratio of variances are statistically significant.

RESULTS

Data were collected from 108 participants for the experiment, the average payout was $5.62 U.S. with a minimum of $4.32 U.S. and a maximum of $6.45 U.S. From the set of mixed-effect logistic regression candidate models, the model with the lowest AIC, Model 1, was selected as the best supported candidate model and used for reported statistical inference from the experiment (Table 3.2). Model 1 included the fixed effects psychological distance, message delivery method, infection risk, and play order, in addition to participant as a random effect. ΔAIC quantifies the loss of information if a different candidate model is used. Neither the social cue experimental variable nor interaction effects were included in the AIC-selected best candidate model. Interaction effects between experimental variables were not found to be significant in any of the candidate models generated.

Model	PD	M	IR	PO	SC	M*IR	M*SC	SC*IR	AIC	ΔAIC
1	X	X	X	X					1495	0
3	X	X	X	X		X			1495	0.280
2	X	X	X	X	X				1498	3.081
7	X	X	X	X	X	X			1499	3.383
4	X	X	X	X		X	X		1502	6.764
5	X	X	X	X				X	1505	10.195
6	X	X	X	X		X		X	1506	10.523

Table 3.2: Candidate Models reordered by AIC value with the best AIC-selected models listed first. Independent variables: Psychological Distance (PD), Message Delivery Method (M), Infection Risk (IR), Play Order (PO) and Social Cue (SC). Interaction terms e.g. Message Delivery Method by Infection Risk are denoted as (M*IR).

The models use a linear combination of the random and fixed effects to obtain logit coefficients that predict the response variable, or the probability (from 0 to 1), that the participant would comply with the shower biosecurity practice. We exponentiated the logit coefficients to generate odds ratios, which were used to evaluate the odds that an individual will opt to use the shower biosecurity practice as opposed to the emergency exit. An odds ratio of 1:1, presented as 1, indicated that there are even odds for either outcome. If an odds ratio confidence interval excluded one, the variable was considered significant. Odds ratios that were greater than 1 indicated that it was significantly more likely that the participant complied with the shower practice than used the emergency exit; ratios below 1 indicated it was more likely the participant used the emergency exit.

Results from the logistic regression quantify the model-predicted probability that the participant will comply with the shower biosecurity practice. Presented as odds ratios in Table 3.3, the first row represents the baseline odds ratio (intercept) associated with the treatment combination of 5% infection risk delivered using a Graphical message. The 66.536 odds ratio (intercept) signifies participants are 66.536 times as

likely to use the shower biosecurity practice as opposed to the emergency exit when provided with the Graphical message at the 5% infection risk level. The rest of the odds ratios in Table 3.3 are compared to the baseline, intercept ratio. For example, participants that received the infection risk information as a Numeric message as opposed to the intercept message delivery method (Graphical), had an odds ratio of 0.095. Therefore, participants receiving a Numeric message were 0.095 times as likely to use the shower biosecurity practice, or 10.526 times as likely to use the emergency exit, than participants receiving a Graphical message. No conclusion can be drawn about how predictor variables affect compliance decisions if the odds ratio confidence interval includes 1. We found significant main effects (Table 3.3, Figure 3.3).

Parameter	Odds ratio	Lower bound	Upper bound	*P*-value
Intercept (Graphical Message, Infection Risk @ 5%)	66.536	22.729	194.774	**<0.001**
Psychological Distance	0.120	0.048	0.299	**<0.001**
Numeric Message	0.095	0.068	0.134	**<0.001**
Infection Risk @ 15%	20.100	12.885	31.353	**<0.001**
Infection Risk @ 1%	0.116	0.081	0.165	**<0.001**
Play Order	0.976	0.949	1.004	0.090

Table 3.3: Results of the selected best fit mixed-effect logistic regression model (Model 1; see Table 3.2). Depicted here are the odds ratios for fixed effects describing relationships with the binary response variable: compliance with the biosecurity practice. Bold values indicate significance at $\alpha = 0.05$.

Main Effects

Average compliance increased with increasing infection risk from 1% (34% compliance) to 5% (60% compliance) to 15% (89% compliance) (Figure 3.3 Left Panel). Compliance was much higher when the infection risk message was delivered as a

Graphical message (73% compliance) vs. as a Numeric message (49% compliance) (Figure 3.3 Right Panel); which was also found in the experiment by Merrill et al. [43].
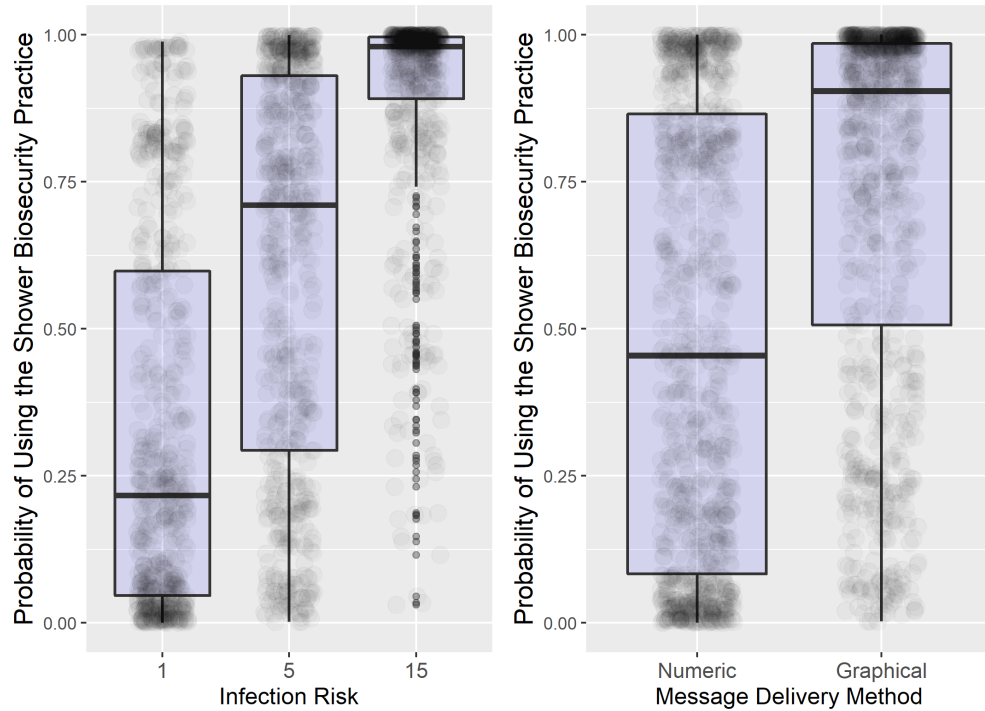
Figure 3.3: Summary results of the main treatment effects. Box-plot of the probability of using the shower biosecurity practice by the main effects, Message Delivery Method and Infection Risk. Lower and upper box boundaries 25th and 75th percentiles, respectively, line inside box median, overlaid on model predicted data values.

INDIVIDUAL DIFFERENCES (H1)

Strategic variability of individual behavior was quantified to determine if the coworker social cues elicited any measurable response. The social cue variable was not included in the AIC-selected best candidate model (Table 3.2). This indicated that individuals did not respond in a consistent or linear way to the coworker demonstrations, so the information gained by including the social cue variable did not explain enough

information to overcome the penalty for inclusion of an additional parameterized variable. While the changes in compliance were not consistent across participants, we sought to discover if the coworker behavior influenced the strategies used by the participants. We identified the strategic variability of individuals by quantifying their change in average compliance when confronted by the different coworker behaviors. Controlling for the effects of the infection risk and message delivery method experimental variables, we calculate individual changes in average compliance between the three social cue treatments: Compliance by Coworker, Non-Compliance by Coworker, and Coworker Control. Specifically, we identify the distributions of average changes in compliance between the three *combinations* of these treatments: Compliance by Coworker vs. Non-Compliance by Coworker, Compliance by Coworker vs. Coworker Control, and Non-Compliance by Coworker vs. Coworker Control (Figure 3.4). This allowed us to test if the strategy space was distributed differently between the social cue treatments.
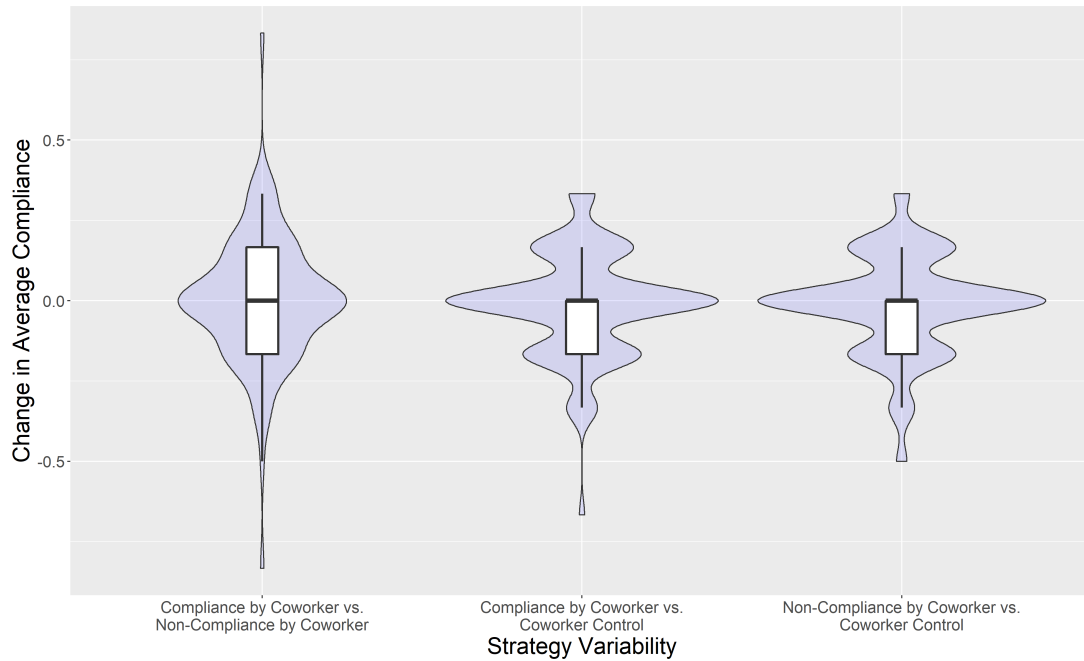
Figure 3.4: Violin with inlaid box-plots of individuals' changes in average compliance between coworker treatments. Lower and upper box boundaries 25th and 75th percentiles, respectively, line inside box median.

Change in average compliance, normalized to span from -1 to 1, represented the degree to which an individual changed their compliance strategy between two types of social cue treatments. Of particular interest was whether participants altered their strategies when observing a coworker exiting the production facility (either following protocols or not) as contrasted with the strategies they employed when the coworker never left the facility. The greatest strategic variability was observed in the changes of compliance between Compliance by Coworker treatments and Non-Compliance by Coworker treatments (Figure 3.4, Left-most violin with inlaid box-plot). In this case, participants with a positive change in compliance (e.g. increasing compliance from a base rate), falling above the y=0 line, were complying with the shower biosecurity practice more often during Compliance by Coworker treatments, when compared with

Non-Compliance by Coworker treatments. However, there were also individuals doing the opposite; participants falling below the $y = 0$ line were complying more often during Non-Compliance by Coworker treatments, when compared with Compliance by Coworker treatments. To evaluate the relative change in variance between the three distributions of compliance strategies (Figure 3.4), we enumerated their ratio of variances (Table 3.4).

		Ratio of variances	
Compliance by Coworker vs. Non-Compliance by Coworker	0.040	1.496	1.680
Compliance by Coworker vs. Coworker Control	0.027	1	1.123
	Variance	0.027	0.024
		Compliance by Coworker vs. Coworker Control	Non-Compliance by Coworker vs. Coworker Control

Table 3.4: *Variance of changes in average compliance between the three combinations of social normative behavior treatments. Ratios of variances quantify the relative change in variance between two distributions.*

To determine if any of the ratio of variances were statistically significant, confidence intervals were calculated using an F-test (Table 3.5). A ratio was determined to be significant if the confidence interval did not include 1. The top two rows of Table 3.5 with significant p-values correspond to the increased variation in individual compliance strategies observed between the two explicit demonstrations (Compliance by Coworker vs. Non-Compliance by Coworker); with respect to compliance strategy variation between either of the two explicit demonstrations and the baseline treatment (Compliance by Coworker vs. Coworker Control, or Non-Compliance by Coworker vs. Coworker Control). This confirmed the hypothesis (H1) that participants would change their behavior using the social cues presented to them. In this case the change

in behavior came in the form of increased strategic variability between either of the two explicit social cues with respect to the control. The third row indicates there was not a significant difference in strategic variability between Compliance by Coworker vs. Coworker Control, and Non-Compliance by Coworker vs. Coworker Control.

Paired Social Cue Treatment Ratio	Ratio of variances	Lower bound	Upper bound	*P*-value
(SC1 vs. SC2) / (SC1 vs. Control)	1.496	1.022	2.190	**0.038**
(SC1 vs. SC2) / (SC2 vs. Control)	1.680	1.148	2.459	**0.008**
(SC1 vs. Control) / (SC2 vs. Control)	1.123	0.767	1.643	0.550

Table 3.5: Results of F-test to calculate ratio of variances (Table 3.4) confidence intervals. SC1 corresponds with Compliance by Coworker treatments, SC2 corresponds with Non-Compliance by Coworker treatments, Control corresponds with Coworker Control treatments. Bold indicates statistical significance, α = 0.05.

Psychological Distance (H2)

The results of the AIC-selected best fitting model from the experiment found the psychological distance effect to be significant with an odds ratio of 0.120 (Table 3.3). This confirms (H2) that temporally-based psychological distancing occurred by indicating that the probability that individuals complied with the shower biosecurity practice increased directly after becoming infected, and the effect decayed with time.

Discussion

The goal of our experiment was to investigate the effects of social cues in a simulated pork production facility. Our study advances knowledge of how heuristics [51], or affective reasoning [48, 49], may impact decisions made under time pressure. We

examined the effects of introducing social cues by testing for disease infection risk, message delivery method, and social cues. Our results unveil the potential impacts of social value orientation [56] by identifying effects of different social cues at the operational level of biosecurity. Social cues were shown to have a significant effect on the degree of strategic variability in compliance behavior (H1). That is, people reacted and changed their behavior when they observed the coworker's biosecurity decision as compared to when the coworker staying in the facility for the entire working day. We also found evidence of a psychological distance effect (H2), an increased likelihood to comply directly after an infection event, an effect that decayed over time.

LIMITATIONS

Statistical sampling methods are often used to study selected groups and strata of society, but they do carry the potential to introduce bias into an experiment when the sampled population has marked differences from the target population. For our study, we assume that results obtained using the online platform Amazon Mechanical Turk can be extrapolated to the target population of swine industry professionals. Amazon Mechanical Turk has been identified as a viable alternative to traditional sampling methods like surveys [61], and is also characterized as a more representative sample of the U.S. national population than either college undergraduate or internet samples in general [60]. Although limited research exists to compare the differences between swine industry workers and the general population, an experiment by Clark et al. [42] did not detect a difference in the distributions of risk behavioral strategies between a sample of online participants and agricultural professionals. Farmers in particular exhibit variable responses to stimuli due to the complexity of the decision-making

process [57]; and because they are operating under a variety of different objective functions a consistent bias is unlikely.

Merrill et al. [43], specifically their second experiment, was the foundation for the experiment within the current study. It is important to acknowledge the differences in the current experiment, both to enhance replicability, as well as to address any potential bias in comparison. In the current experiment with respect to the work of Merrill et al. [43]: the linguistic message delivery treatment was removed, the social cue experimental variable was introduced, the appearance of the threat gauge was adjusted, the experiment had six fewer rounds, and each round was approximately 10 seconds longer. Although the look of the threat gauge was modified, the relatively consistent use of verbal demarcations (Low, Medium, High), should preserve response to the graphical message across experiments.

INFECTION RISK

Infection risk has been determined to be the main driver of behavior both in this current experiment, as well as the work of Merrill et al. [43]. In the study by Merrill et al. [43], participants were exposed to infection risk treatments of very low (1%), low (5%), medium (15%) and high (25%). In the current experiment, the high infection risk treatment with a 25% chance of infection was omitted because Merrill et al. [43] observed near ubiquitous compliance at 25%, and we wanted to increase our ability to detect differences in other signals, such as the social cue experimental variable. Therefore, participants were only confronted with the very low (1%), low (5%), and medium (15%) infection risk treatments. The change in risk perception that could be attributed to the removal of the high infection risk treatment can be seen when

comparing average compliance in the current experiment, with average compliance in analogous treatments from the experiment of Merrill et al. [43] (Table 3.6). Here analogous refers to message delivery method treatments in the experiment conducted by Merrill et al. [43] that were replicated in the current experiment, numeric with certainty and graphical with uncertainty; with message delivery method treatments that were not replicated, e.g. numeric with uncertainty, omitted from comparison.

Infection risk	Merrill et al. frequency of compliance (%)	Current study frequency of compliance (%)	Δ frequency of compliance (%)
1	23.6	33.6	10.0
5	56.0	60.3	4.3
15	80.9	88.7	7.8
Average	53.5	60.9	7.4

Table 3.6: *Change in observed frequency of use of the shower biosecurity practice between original analogous treatments and current experiment treatments by infection risk. Δ frequency of compliance quantifies the change between the original and current average compliance, positive values correspond to an increase in the current study average compliance.*

At analogous treatments in the experiment of Merrill et al. [43], average compliance increased from 23.6% at very low infection risk, to 56.0% at low risk, to 80.9% at medium risk, to 91.0

Individual Differences (H1)

By introducing the social cue variable into the experiment design, we hypothesized that compliance behavior could be shifted by an implicit suggestion. In this case, the social cue came from a coworker present in the facility with the participant. We observed increased strategic variability by participants in their response to compliance or non-compliance by the coworker, with respect to the control where the coworker

never exited. The behaviors of the automated coworker could have changed how a participant perceived what is commonly done in the facility. Although the automated coworker had no effect on the actual outcome of the round, we expected the subtle hint of how the coworker handled the compliance decision to be internalized in the decision-making process of participants. Humans are known to let feelings and mental shortcuts guide the decision-making process, especially when under time pressure. [48, 49, 51]

In our examination of the variability in the participants' strategy space, we established that individuals displayed significantly more variation in compliance strategy when responding to two different explicit social behaviors: Compliance by Coworker and Non-Compliance by Coworker. The social value orientation of participants, associated with varying tendencies to cooperate or compete with interaction partners, likely played a role in observed differences in strategic variability. [56] These differences could also be attributed to an mirroring or mimicry effect; imitation based on social performance cues has been observed, although in this experiment there was no effect on performance or payout related to the cue. [58]This aligns with research identifying the impact of social cues on behavioral flexibility, leading to novel behavioral patterns. [59] Quantified as variance, or ratio of variances, having a significant increase indicates individuals were changing, but not in a consistent way. One explanation is that individuals are known to be motivated by a wide range of factors and internal logic. [57] The complex decision-making process in each individual may lead to varying degrees of "follower" or "anti-follower" tendencies. If a worker in a facility sees their associate breaking the rules, are they more likely to pick up the slack and "cover for them", or are they more likely to feel it is justified for them to break the

rules as well? One player who shared their strategy after the experiment stated they would always use the shower biosecurity practice when the coworker was exiting the emergency door to compensate/offset their coworker's bad behavior. They felt more inclined to shower in order to protect the farm from disease.

Identifying that social cues influence compliance strategy at the operational level is a novel conclusion and lends credence and support to future studies of worker culture. We recommend follow-up experiments with more explicit social cues, e.g., where a coworker demonstration is paired with an associated explanation of behavior communicated via a text bubble. For example, how would pairing the coworker demonstration of compliance within this experiment with an explicit verbal explanation from the coworker (e.g. "Showering takes time, but we are all in this together.") affect behavior of participants? Another avenue to further research effects of social cues on operational biosecurity compliance is to replicate the current experiment in a virtual-reality based swine production facility, in order to test if the results are consistent in a more immersive environment. Continuing to study how social cues can increase compliance, or induce non-compliance, will be valuable for farm management, training, and monitoring.

PSYCHOLOGICAL DISTANCE (H2)

In our experiment, individuals who had recently incurred a facility infection were far more likely to comply with the shower biosecurity practice. Over time, mental construal of events are expected to become more abstract and low-level, decreasing their influence on future decisions. [30] In the context of this experiment, the increased likelihood to comply directly after an infection event wore off over time. This tem-

porally based psychological distancing [29, 31] was similarly observed by Merrill et al. [43], where individuals who had just been infected were twice as likely to comply as those who had never been infected. Our results therefore suggest that it is important to acknowledge that temporal psychological distancing will influence workers and the decrease in observed compliance should be combatted such as with biosecurity trainings that are reinforced frequently.

Conclusion

The human behavioral component of animal biosecurity is not well understood, but worker decisions at the operational level have direct economic and sociological consequences when an outbreak occurs. [19,28] For both managers and workers in the swine industry, understanding how social cues are affecting compliance with existing biosecurity protocols is critical. This study demonstrates the ability to test hypotheses about human behavioral responses to social cues using experimental game simulations. We examined how a coworker demonstration can impact the complex mental process occurring when participants chose to comply with the simulated shower biosecurity practice in the experiment. We tested the effect of an implicit social cue on decision-making; an attempt to subtly recreate the worker culture in a facility. [46] While it is well known that workers do not always comply with operational-level protocols [28], there is no existing data that shows how social cues may be impacting these biosecurity lapses. These findings outline the significant variability in how people change their behavior in response to those around them, even when others' actions have no repercussions or effect on monetary payout.

To summarize and extend these findings to be applicable at the facility level, this work identifies that social cues will result in different compliance strategies amongst workers in a swine facility. Behavior of others, even if not directly impacting the worker, will impact their decision making. Providing this feedback from the operational level can inform tactical and strategic decision makers as they implement biosecurity protocols whose efficacy will depend on workers' willingness to comply, and attempt to create a workplace culture of compliance.Humans are extremely complex, and while we acknowledge that there is no blanket solution to increasing worker compliance, we believe small changes can have an impact on the system level. Advancing knowledge and understanding of human behavioral components of animal biosecurity has vast potential to increase worker and animal welfare, shifting the industry towards disease-resiliency.

CHAPTER 4

CHARACTERIZATION OF GENE

REGULATION IN ENDOTHELIAL

CELLS THAT ARE PROGRAMMED TO

HUMAN PLURIPOTENT STEM CELLS

AND DIFFERENTIATED TO

ENDOTHELIAL OR NEURONAL CELLS

ABSTRACT

Opportunities for increased health benefits exist through the study of stem cells, which have the potential to develop into many types of cells in the body. Here we develop and implement data analysis to examine gene expression in various types of cells. Endothelial cells line the interior surface of blood and lymphatic vessels and play a crucial role controlling passage of materials, clotting, formation of blood vessels, blood pressure, as well as other physiological and pathophysiological functions. Neurodegenerative diseases like Parkinson's and Alzheimer's are characterized by progressive loss of structure or function of neurons. Human pluripotent stem cells, derived from embryos or fetal tissue, are providing new opportunities to understand changes in gene regulation with applications for modeling and treatment in the realms of neurodegeneration and cardiovascular disease. Induced pluripotent stem cells (iPs) were generated from human umbilical vein endothelial cells (HU-VEC), and then differentiated back into endothelial cells (EC-Diff) as well as into neuronal cells (Nn-Diff). Changes in gene differentiation were grouped based on two gene ontological categories: biological process and cellular component. Transcriptional regulation changes, angiogenic upregulation of EC-Diff, and MAPK pathway upregulation for Nn-Diff are established in biological process gene expression patterns. The density of significant genes in the extracellular region and nucleus as well as a unique expression pattern in the membrane are established in cellular component gene expression patterns. We provide evidence that differentiation of endothelial and neuronal cells is related to changes in genetic expression; and that profiles of changes in genetic expression differ by gene functional categories. Our sequential data analysis, or bioinformatics pipeline, is implemented to identify profiles of gene expression changes grouped by ontology. The results presented here will empower biologists to generate novel conclusions about the function of stem cells, with applications for regenerative medicine.

INTRODUCTION

Millions of people suffer from vascular diseases, related to the vast network of blood vessels in the human body. Endothelial cells compose the endothelium, the interior surface of blood and lymphatic vessels. These cells are vital in vascular permeability, regeneration of tissue, as well as forming new blood vessels from existing blood vessels (angiogenesis). With wide-ranging importance in various physiological functions including innate and adaptive immunity, the endothelium is related to many aspects of pathophysiology. This integrated system transcends clinical disciplines due to blood vessel traversal of all tissues. [70] At the clinical and laboratory level, there is untapped potential to better understand the endothelium and designate it as a therapeutic target. [71] Mechanistic understanding of endothelial-specific gene expression has value to shed light on healthy function as well as disease-related dysfunction of endothelial cells. The increased availability of induced pluripotent stem cells for experiments has allowed the characterization of generated endothelial gene expression patterns and overall expression profiles of endothelial specific genes. [72–74]

Effective therapies have yet to be developed for neurodegenerative diseases such as multiple sclerosis or Huntington's disease. Similar to the analysis confounding heterogeneity in endothelial studies, limitations of animal models aimed at studying human disease have impeded development of effective treatments for diseases affecting neurons. With the generation of induced pluripotent stem cells, a crucial step has been taken towards developing cellular function disease models that advance research of neurodegenerative diseases. [10] In this study, gene expression data was analyzed from an experiment in which induced pluripotent stem cells were generated from human

umbilical vein endothelial cells and then differentiated back into endothelial cells as well as neuronal cells. Here we hypothesized that differentiation of endothelial and neuronal cells is related to changes in genetic expression. Gene regulation was accessed for significance based on changes in expression. Patterns in expression were explored based on two categories of gene ontology: biological process and cellular component. A refined bioinformatics pipeline was developed in collaboration with biologists who have domain expertise in the area of genetic expression. The implementation and resulting output from the pipeline demonstrates that profiles of changes in genetic expression differ by gene functional categories. Future work to confirm the connection between changes in gene expression and biological pathways has the potential to advance the frontiers of regenerative medicine.

METHODS

INITIAL ANALYSIS

The initial dataset contained 49400 observations for 19929 unique genes. Each observation had multiple gene expression values corresponding to induced pluripotent cells, differentiated endothelial cells, and differentiated neuronal cells. Some genes had more observations than others, so for visualization purposes, the mean expression over all observations of a gene was used.
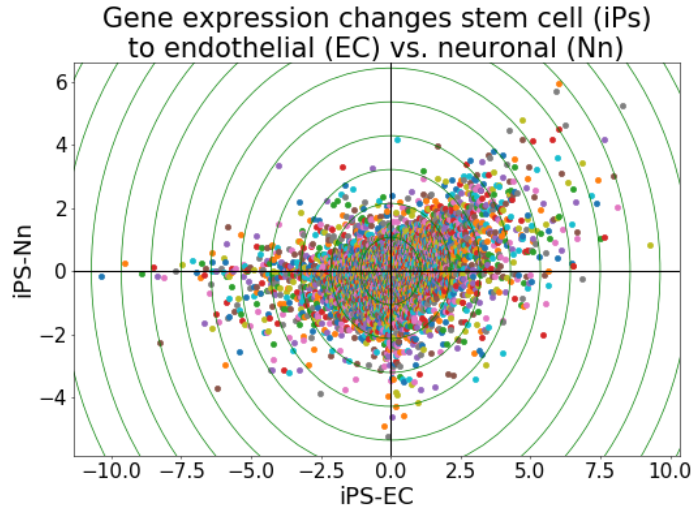
Figure 4.1: Scatterplot of average gene expressions. The x-axis labelled 'iPS-EC' indicates gene expression in differentiated endothelial cells subtracted from gene expression in induced pluripotent cells. The y-axis labelled 'iPS-Nn' indicates gene expression in differentiated neuronal cells subtracted from gene expression in induced pluripotent cells.

Genes with significant differential expression were selected using the 'fcros' package in R in order to estimate the likelihood a significant change in expression occurred. [12, 75] This method detects differentially expressed genes based on the fold change rank, which is a ratio of means from control to test samples. The control expression used was gene expression from induced pluripotent cells. The two test expressions used were differentiated neuronal, and differentiated endothelial cells, respectively. Genes were considered significantly upregulated if their adjusted p-value was above 0.9577 and considered significantly downregulated if their adjusted p-value was below 0.0423. In addition, any gene with an absolute expression change greater than three was included as significant. Note that these thresholds, which were used based on visual inspection and default parameterization in the 'fcros' package, were adjusted in the following pipeline implementation based on consultation with domain experts. At this stage a total of 772 genes were identified as significant from the ini-

tial dataset. Potentially significant genes were grouped by two ontological categories provided in the dataset: biological process and cellular component. Groups were selected for visualization based on multiple factors. For biological processes, eight groups were included due to prominence of clustering, size of group, and relevance in literature. For the cellular component gene ontology grouping, the largest groups had the majority of significant genes. Therefore, the ten largest groups were selected for visualization.

Pipeline Implementation

Detailed documentation for the pipeline, which included hosting of Python [76] and R [64] scripts in an online repository, facilitated testing and allowed for users' input as the pipeline was built. In the initial analysis, genes were considered to have a significant change in expression if their fold rank change adjusted p-value was above 0.9577 or below 0.0423, e.g. $\alpha = 0.0423$. In implementation of the pipeline, a slightly less restrictive threshold is utilized, $\alpha = 0.05$. In addition, genes were included as significant if their mean change in genetic expression from induced pluripotent stem cell to neuronal or endothelial cell was greater than 1.5 or less than -1.5. It was determined that based on the specified pipeline thresholds, β-catenin (CTNNB1), a gene known to be important in the maintenance of stem cells [77], would not be included. Here we implement a third threshold, based on the fold rank change of β-catenin ($\alpha = 0.218$), to identify genes in the ontological groups we are interested in that may not have met the stricter criteria.

Having identified significant and potentially significant ($\alpha \leq 0.218$) genes using the pipeline, we then subset the genes based on six gene ontology [13] classifications

which were identified as relevant in maintenance and/or differentiation of stem cells: Signaling, Transcription, Epigenetic, Cell Adhesion, and Extracellular Matrix Protein. Having subset the dataset of genes, we then calculate the proportion falling into each quadrant of change in expression (see Figures 4.2 and 4.3), to determine the general profiles of expression change.

Results

Initial Output
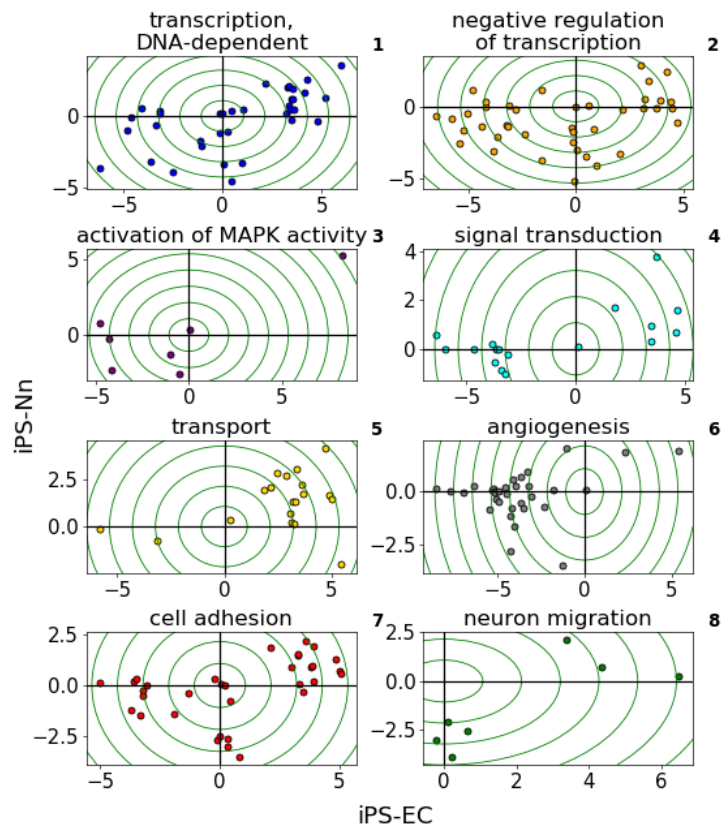
Figure 4.2: Average gene expression patterns grouped by biological process.
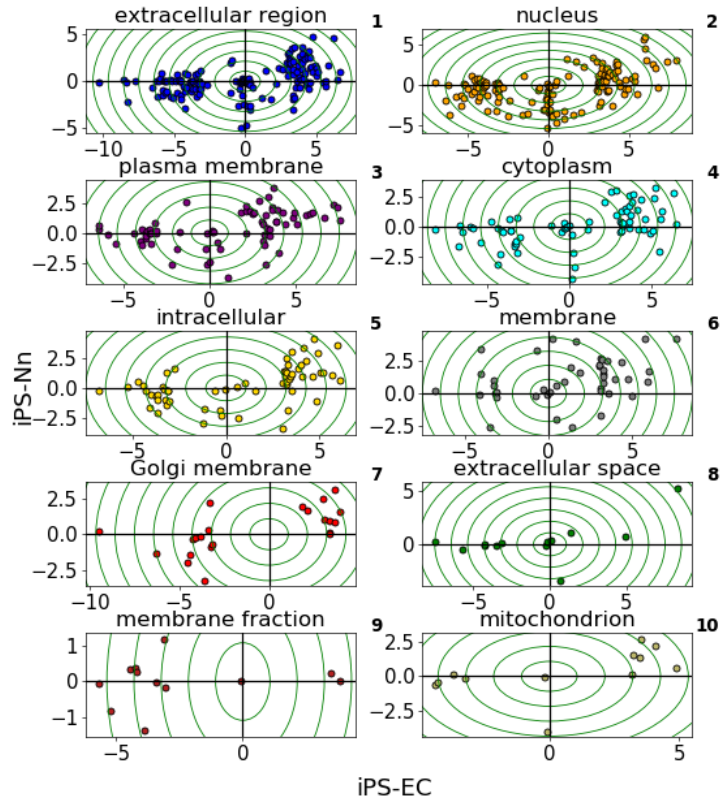
Figure 4.3: Average gene expression patterns grouped by cellular component.

Pipeline Output
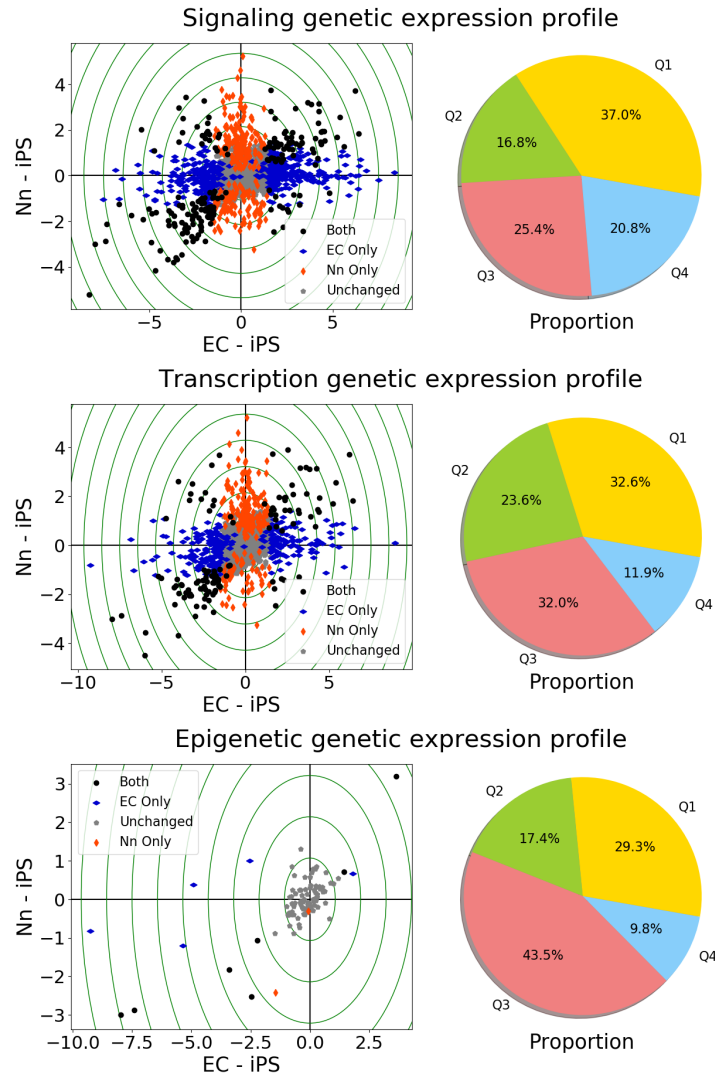
Figure 4.4: Average gene expression grouped by gene ontological subsets: Signaling, Transcription, and Epigenetic; pie charts represent proportion within the subset in each quadrant. Black dots identified as 'Both' represent genes meeting the criteria for fold rank change significance $\alpha = 0.05$, or having absolute mean change greater than 1.5, in differentiation to both endothelial and neuronal cells. Blue diamonds represent genes meeting either of the significance conditions for only endothelial cells, marked 'EC'; red diamonds represent genes meeting either of the significance conditions for only neuronal cells, marked 'Nn'. Genes that met neither condition are plotted as gray pentagons and marked 'Unchanged'; this indicates they were included by having a fold change rank significance based on a threshold of $\alpha = 0.218$.

Figure 4.5: Average gene expression grouped by gene ontological subsets: Metabolic, Cell Adhesion, and Extracellular Matrix Protein; pie charts represent proportion within the subset in each quadrant. Black dots identified as 'Both' represent genes meeting the criteria for fold rank change significance $\alpha = 0.05$, or having absolute mean change greater than 1.5, in differentiation to both endothelial and neuronal cells. Blue diamonds represent genes meeting either of the significance conditions for only endothelial cells, marked 'EC'; red diamonds represent genes meeting either of the significance conditions for only neuronal cells, marked 'Nn'. Genes that met neither condition are plotted as gray pentagons and marked 'Unchanged'; this indicates they were included by having a fold change rank significance based on a threshold of $\alpha = 0.218$.

Discussion

Biological Processes

In the initial analysis, the two largest numbers of significant gene changes related to biological processes were from 'transcription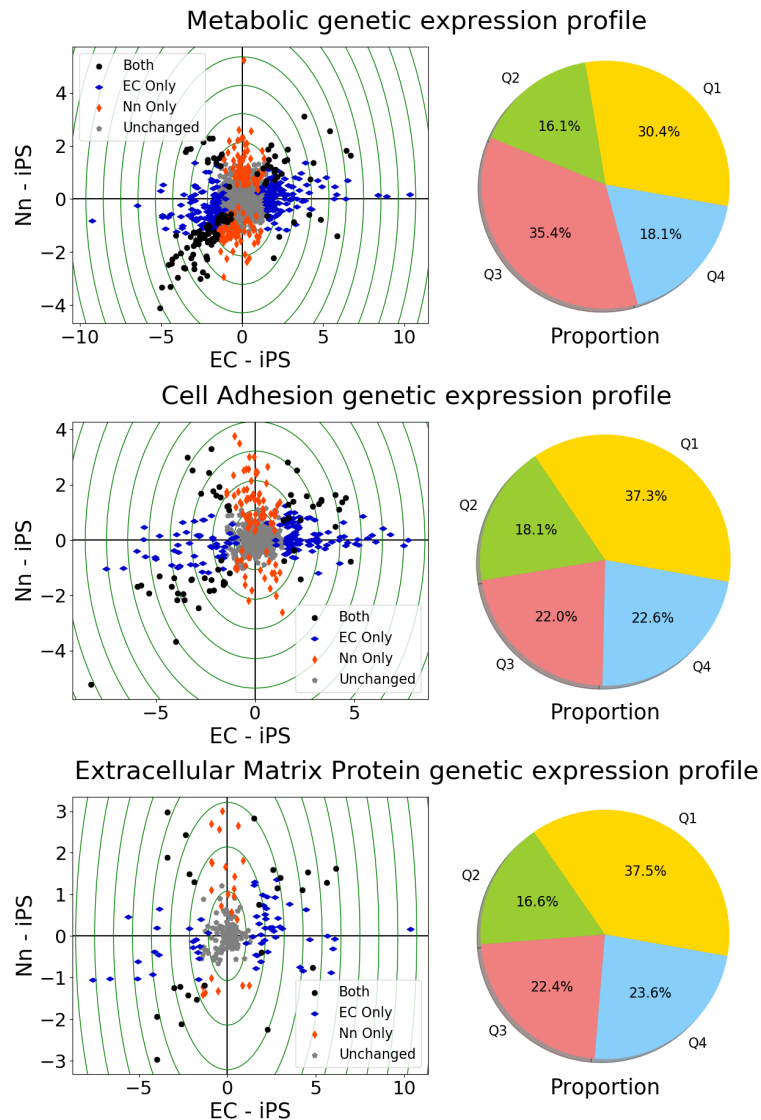, DNA-dependent' and 'negative regulation of transcription from RNA polymerase II promoter' genes. These groups had similar patterns of expression seen in the first two subplots of Figure 4.2. For average expression values falling below the x-axis and to the left of the y-axis, the third quadrant, this indicates upregulation of gene expression in both differentiated neuronal and endothelial cells. Conversely, downregulation of genes in both differentiated neuronal and endothelial cells results in observations in the first quadrant. Transcription is seen here to be accompanied by various cohorts of regulation. Previously, endothelial differentiation has been used to study changes in transcriptional regulation associated with revoking and establishing endothelial cells. [72]

Mitrogen-activated protein kinases (MAPKs) are a type of protein enzyme that catalyze the transfer of a phosphate group from ATP to a specified molecule. These enzymes are involved in fundamental cellular processes, and have been shown to be crucial in differentiation of endothelial cells. [78] Certain genes are known to activate the MAPK pathways, and the third subplot in Figure 4.2 supports this conclusion by showing upregulation in those genes in differentiated endothelial cells. Signal transduction is the process of signal transmission through a cell and contributes to many basic mechanisms such as cell growth or communication. Signal transduction genes in the fourth subplot of Figure 4.2 show both up and downregulation in differentiated

endothelial cells. Transport genes shown in the fifth subplot show downregulation in both types of differentiated cells. Previous work has shown the capabilities of endothelial cells to be a source for vascular repair [72], and the sixth subplot supports this notion by showing angiogenesis genes to be upregulated in differentiated endothelial cells. Cell adhesion is the process of adjacent cell interaction facilitated by specialized protein complexes. The pattern of gene expression related to cell adhesion in the seventh subplot of Figure 4.2 is similar to that of transcription in the first two subplots. Neuron migration is part of nervous system development, a method by which neurons are brought together to interact properly. The eighth subplot showed some neuron migration genes were upregulated in differentiated neuronal cells, while some were downregulated in differentiated endothelial cells, meaning the expression of these genes is not consistent across cell types.

Cellular Component

Grouping by cellular component, seen in Figure 4.3, resulted in higher density plots for gene expression in domains like the extracellular region and nucleus. This could represent that genes identified in these components are more likely to be significant, or simply that more testing has been done on genes in these areas. Some of the most consistent extreme expression changes are from upregulation of genes in the extracellular region, seen in the first subplot of Figure 4.3. The term extracellular region is intended to annotate gene products that are not attached to the cell surface or internal to the cell. In differentiated neuronal cells, the nucleus has a high number of upregulated genes seen in the second subplot; while the membrane had a large number of downregulated genes seen in the sixth subplot. The membrane related gene

expression pattern is interesting because multiple observations are seen in quadrant two. Observations in the second quadrant, corresponding to upregulation in differentiated endothelial cells and downregulation in differentiated neuronal cells, are rare. The plasma membrane forms the external boundary of the cytoplasm of a cell. It is composed of a phospholipid bilayer with embedded proteins and functions to protect the cell from its surroundings. The expression pattern of plasma membrane genes seen in the third subplot of Figure 4.3 is similar to that of cytoplasm and intracellular related genes in subplots four and five. Cytoplasm is the material within the cell that gives the cell its shape, so it is logical cytoplasm-related gene expression would be associated with intracellular gene expression.

The Golgi membrane functions as a protein processor and transports lipids. Genes associated with the Golgi membrane in the seventh subplot of Figure 4.3 show significant up regulation in differentiated endothelial cells. This trend is seen again in gene expression related to the membrane fraction, or fraction of cells prepared by disruptive biochemical methods i.e. plasma, pictured in the ninth subplot. The mitochondrion is related to the biochemical processes of respiration and energy production. As the sample size of groups decreases, patterns are more difficult to infer due to sparse data. This is seen in mitochondrion gene expression in the tenth subplot of Figure 4.3; and also in neuron migration gene expression in the eighth subplot of Figure 4.2.

Pipeline Reporting

The significant and potentially significant genes identified via the pipeline are visualized in Figures 4.4 and 4.5. Subplots are grouped by gene ontological classification. For each subset, the proportion of genes falling into each quadrant (1, 2, 3, 4) are

visualized using a pie chart. For example, the transcription subset in Figure 4.4 had the largest proportion of genes that were downregulated in neuronal cells, and upregulated in endothelial cells, with 23.6% falling into Quadrant 2. This marks a departure from the initial output, so here we define how the x- and y-axis in the pipeline output were altered from the initial analysis, where they previously denoted iPs-Nn and iPs-EC, respectively. Here it was determined that placing genes that are upregulated in both EC and Nn in Quadrant 1 (as opposed to Quadrant 4) would provide a more intuitive understanding of the changes in regulation. Thus, the subtraction was reversed to be Nn-iPs and EC-iPs, respectively.

The scatterplot pipeline outputs augment the initial analysis by using color and shape to indicate which criteria for significance a gene may have met. Recall that there were two criteria for significance in the pipeline as opposed to the initial analysis, fold rank change significance with $\alpha = 0.05$, and absolute mean change greater than 1.5. If a gene was determined to meet both conditions, it is visualized as a black point in Figures 4.4 and 4.5. Genes that were determined to be significant in their expression change for either endothelial or neuronal cells, but not both, are plotted as blue or red diamonds, respectively. To delineate genes that met neither of the conditions for significance but did pass the third less restrictive criteria, based on the fold rank change value of β-catenin ($\alpha = 0.218$), a gray pentagon was utilized.

The results of the pipeline output confirm the hypothesis that gene expression is related to differentiation of stem cells into endothelial and neuronal cells, and that profiles of gene expression changes can be grouped by the biological function of genes. This work is poised to play a critical role in developing novel conclusions about genetic expression profiles and biological pathway activation in differentiation of stem cells.

Although Figures 4.4 and 4.5 have yet to be interpreted, we propose a future study that utilizes the analysis visualized here to confirm activation of certain biological pathways. The application of our pipeline within this work is novel, and in addition to allowing us to identify potentially significant genes, it has great potential for future hypothesis generation. Tables which list the genes visualized in each figure, as well as analysis source code, are available upon request to the author.

Conclusion

Efficient differentiation of induced pluripotent cells to endothelial and neuronal cells requires increased mechanistic understanding of gene expression. This area of research has the potential to contribute in developing novel treatments for patients suffering from cardiovascular or neurodegenerative diseases. The visualization and characterization undertaken in the context of this study demonstrate a process to distill gene expression data into more salient visual pieces. Grouping by gene ontology has benefits in capturing the overall picture of different patterns in gene expression. Delineating profiles of gene expression and clearly demarcating the significance of genes through their expression changes is beneficial to biologists. The implementation of our bioinformatics pipeline to analyze data sequentially was vital in augmenting the initial analysis and output of the experiment; enabling the generation of results that have the potential to advance the frontier of regenerative medicine.

CHAPTER 5

CONCLUSION

Within this work we have defined complex systems research as the study of systems with highly variable behavior, which contain many mutually interacting parts. While there are many systems that meet this definition in the natural world, the tools utilized in uncovering the underlying mechanisms of such systems remain consistent when applied across domains. Experience with complex systems and a combination of computer science, statistics, and mathematics was necessary in order to reach the conclusions presented in this work, unveiling emergent behavior that would have been difficult or impossible to extract otherwise. To discuss some of the broader methods of this work that generalize well to different applications of complex systems research, as well as present more specific details of implementation, an overview of methods was presented.

Emphasis on design patterns and software development practices is integral to successfully implementing any application-specific methods. A thorough understanding of design patterns, which encompass a reoccurring problem and the tradeoffs of possible solutions, enables efficiency and reusability in software development. Embracing

many iterations of rapid development cycles facilitates collaboration and ensures a more robust path to deploying working platforms. Framing products of research (e.g. journal articles) as intermediate outputs enables better long-term strategizing to comprehensibly achieve project goals. In addition to design patterns, more specific aspects of our methods were discussed, including user interface design and bioinformatics pipelines.

User interfaces facilitate interaction with technology by utilizing objects that aid users in creating mental models of functionality. The hierarchical nature of constructing such an interface was discussed and visualized, in addition to the different types of canvas which compose a single frame within a simulation. Functionality of various components was discussed through examples (e.g. virtual coins) which illustrate the complexity of simulation design. Presenting these specifications within the framework of complex systems research is novel; and could allow researchers to take advantage of interface design in areas where its usage is less common (e.g. animal biosecurity).

The background of designing a bioinformatics pipeline to handle sequential transformation of data was introduced, as well as an explanation of the gene ontology classification system. Our specific pipeline was visualized, which can validate existing pipelines as well as be used independently to analyze experiments producing datasets that meet our input criterion. Having already completed many iterations of testing with domain experts, the pipeline presented here incorporates elements of software design, as well as considerations important to biologists. This work advances practices within a more recently established research area, the study of regenerative medicine using induced pluripotent stem cells, which can assist in future pipeline design.

After providing an overview of methods, we then presented two applications of complex systems research in the domains of animal biosecurity and genetic expression. A serious gaming methodology was utilized to study the effects of social cues on operational biosecurity compliance. We provided evidence that participants in the experiment changed their behavior when observing a social cue, despite the cue having no effect on participants' payouts. These results have benefits for farm management and monitoring, as well as providing an impetus for future study of worker culture. Additionally, the analysis of variance components within Chapter 3 exemplifies an application of statistics to the problem space that would not have been possible without approaching the question from a complex systems perspective.

Following the animal biosecurity application, we then utilized the methods of complex systems science to characterize gene regulation. An experimental dataset was analyzed using our bioinformatics pipeline to develop novel outputs that would not be possible without effective algorithm design and scripting. Here we provided initial analysis and discussion related to various biological processes and cellular components, as well as results produced in collaboration with biologists that have yet to be interpreted. Future conclusions based on the pipeline outputs presented here will unveil previously unknown details of stem cell function, with potential real-world applicability in the field of regenerative medicine.

To summarize the uniqueness of the methods within this work, the applications described here necessitated an advanced understanding of a multitude of subject areas. This represents an approach that emphasizes rapidly acquiring knowledge of a novel domain area that can benefit from complex systems tools. Effectively solving the problems in this work required a careful choice of which tool to apply, despite the

problems being crafted in such a way that choosing a tool was an extremely open-ended process. Having summarized the conclusions of the work presented thus far, we now describe areas of ongoing future work in the domains of animal biosecurity and water quality.

CHAPTER 6

FUTURE WORK

ANIMAL BIOSECURITY

In continuing to work towards the goal of increasing disease resiliency in animal production chains, multiple avenues are currently under exploration. There is considerable interest in replicating the results of Merrill et al. [43] using different languages (e.g. Spanish or Chinese), and immersive virtual reality environments. Cultural differences could be explored by comparing the effects of messaging or social cues across populations. In addition, novel hypothesis can be tested regarding potential effects of virtual reality on willingness to comply with operational biosecurity practices.

Figure 6.1: Virtual reality development environment depicting a simulated pork production facility.

Revisiting the strategic, tactical, and operational level framework which assists in research of animal biosecurity, there is ongoing work to study decision-making at the tactical farm management level. Specifically, how will individuals allocate resources to prevent diseases in a network where multiple farms fall under their control. We hope to understand this optimization problem by developing simulations which allow individuals to take the role of farm manager and defend their facilities. In addition, it is highly relevant that vulnerabilities in a system are well-understood. We aim to add sequential functionality to simulations; in which certain participants defend a system, followed by other participants who can attack the same system i.e. bioterrorism. This work has generated interest in the field of intelligent enemy research, with potential application for defense of strategic interests.
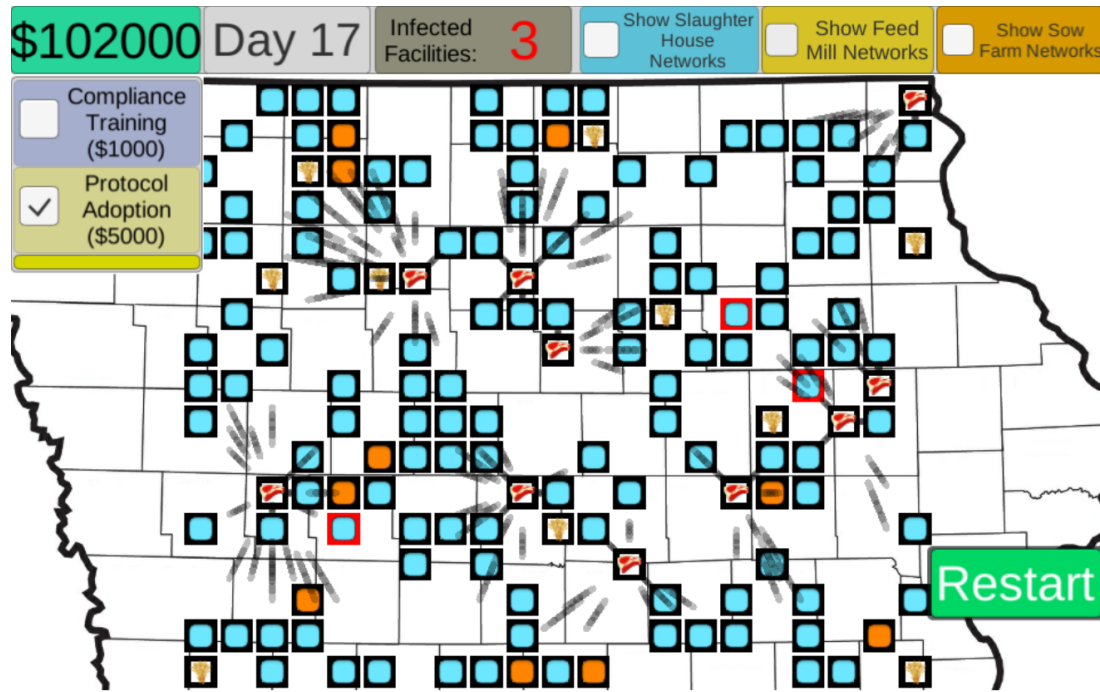
Figure 6.2: Agricultural production system network simulation interface.

WATER QUALITY

Understanding how pollution affects our lakes and rivers, and where it comes from, is an extremely complicated problem. Research is currently underway to shed light on how we can shift agricultural systems to better protect natural resources. Human behavioral factors are known to have an effect in these systems. For example, it is commonly believed that applying manure before a rainstorm is the best way to help it soak in and promote crop growth; however, research shows that this leads to more runoff that damages our waterways. Certain land management practices have been identified as effective ways to decrease pollution from farming, such as vegetative buffer strips which absorb nutrients that would otherwise degrade water quality. Here

we seek to develop a simulation that can help policy makers understand the effects of flood and drought on the installation of best management practices.
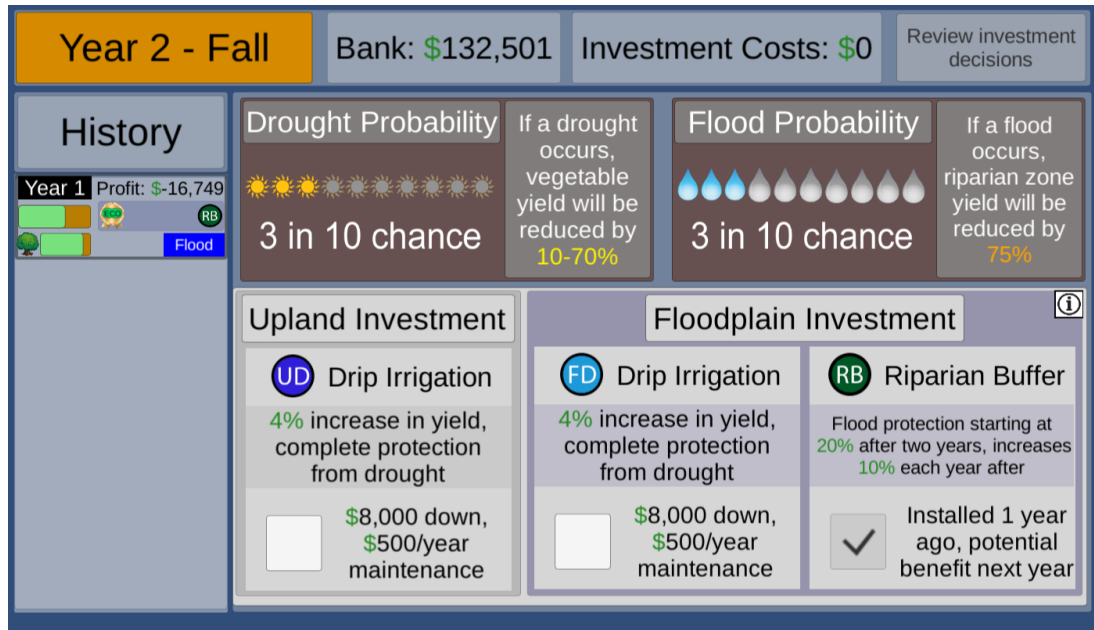

Figure 6.3: Farm management user interface testing effects of flood and drought predictions on investments into best management practices.

Bibliography

[1] Yaneer Bar-yam. Dynamics of Complex Systems. 1997.

[2] Cosma Rohilla Shalizi. Methods and techniques of complex systems science: An overview. In *Complex systems science in biomedicine*, pages 33–114. Springer, 2006.

[3] Nenad Medvidovic and Richard N Taylor. Software architecture: foundations, theory, and practice. In *Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering-Volume 2*, pages 471–472. ACM, 2010.

[4] Erich Gamma. *Design patterns: elements of reusable object-oriented software.* Pearson Education India, 1995.

[5] Kent Beck, Mike Beedle, Arie Van Bennekum, Alistair Cockburn, Ward Cunningham, Martin Fowler, James Grenning, Jim Highsmith, Andrew Hunt, Ron Jeffries, et al. Manifesto for agile software development. 2001.

[6] Ben Shneiderman, Catherine Plaisant, Maxine Cohen, Steven Jacobs, Niklas Elmqvist, and Nicholas Diakopoulos. *Designing the user interface: strategies for effective human-computer interaction.* Pearson, 2016.

[7] Mary Beth Rosson and John M Carroll. Scenario-based design. In *Human-computer interaction*, pages 161–180. CRC Press, 2009.

[8] Anders Hejlsberg, Scott Wiltamuth, and Peter Golde. *C# language specification.* Addison-Wesley Longman Publishing Co., Inc., 2003.

[9] Jeremy Leipzig. A review of bioinformatic pipeline frameworks. *Briefings in bioinformatics*, 18(3):530–536, 2017.

[10] Yohannes Haile, Maryam Nakhaei-Nejad, Paul A Boakye, Glen Baker, Peter A Smith, Allan G Murray, Fabrizio Giuliani, and Nadia Jahroudi. Reprogramming of huvecs into induced pluripotent stem cells (hipscs), generation and characterization of hipsc-derived neurons and astrocytes. *PLoS One*, 10(3):e0119617, 2015.

[11] Maria A Lagarkova, Maria V Shutova, Alexandra N Bogomazova, Ekaterina M Vassina, Evgeny A Glazov, Ping Zhang, Albert A Rizvanov, Ilya V Chestkov, and Sergey L Kiselev. Induction of pluripotency in human endothelial cells resets epigenetic profile on genome scale. *Cell cycle*, 9(5):937–946, 2010.

[12] Doulaye Dembélé. Analysis of high-throughput biological data using their rank values. *Statistical methods in medical research*, 28(8):2276–2291, 2019.

[13] Gene Ontology Consortium. The gene ontology (go) database and informatics resource. *Nucleic acids research*, 32(suppl_1):D258–D261, 2004.

[14] Christopher Pudenz, Lee Schulz, and Glynn Tonsor. Biosecurity and Health Management by US Pork Producers - 2017 Survey Summary. 2017.

[15] National Pork Board. Annual PRRS Costs Fall $83.3 Million - Productivity gains blunt the impact of PRRS on the U.S. herd. https://www.pork.org/blog/annual-prrs-costs-fall-83-3-million-productivity-gains-blunt-impact-prrs-u-s-herd/, Jun 2017.

[16] Philip Paarlberg. Updated Estimated Economic Welfare Impacts Of Porcine Epidemic Diarrhea Virus (Pedv). *IDEAS Working Paper Series from RePEc*, 2014.

[17] PJ Sanchez-Cordon, M Montoya, AL Reis, and LK Dixon. African swine fever: A re-emerging viral disease threatening the global pig industry. *The Veterinary Journal*, 233:41–48, 2018.

[18] Ann M Cummins, Nicole J Olynk Widmar, Candace C Croney, and Joan R Fulton. Understanding consumer pork attribute preferences. *Theoretical Economics Letters*, 6(02):166, 2016.

[19] Caroline Ritter, Jolanda Jansen, Steven Roche, David F Kelton, Cindy L Adams, Karin Orsel, Ron J Erskine, Geart Benedictus, Theo JGM Lam, and Herman W Barkema. Invited review: Determinants of farmers' adoption of management-based strategies for infectious disease prevention and control. *Journal of Dairy Science*, 100(5):3329–3347, 2017.

[20] Arata Hidano, Gareth Enticott, Robert M. Christley, and M. Carolyn Gates. Modeling Dynamic Human Behavioral Changes in Animal Disease Models: Challenges and Opportunities for Addressing Bias. *Frontiers in Veterinary Science*, 5:137, 2018.

[21] Aditi Mankad. Psychological influences on biosecurity control and farmer decision-making. A review. *Agronomy for Sustainable Development*, 36(2):1–14, 2016.

[22] Günter Schmidt and Wilbert E Wilhelm. Strategic, tactical and operational decisions in multi-national logistics networks: a review and discussion of modelling issues. *International Journal of Production Research*, 38(7):1501–1523, 2000.

[23] Derk Loorbach. Transition Management for Sustainable Development: A Prescriptive, Complexity-Based Governance Framework. *Governance*, 23(1):161–183, 2010.

[24] Victor M. Andres and Rob H. Davies. Biosecurity Measures to Control Salmonella and Other Infectious Agents in Pig Farms: A Review. *Comprehensive Reviews in Food Science and Food Safety*, 14(4):317–335, 2015.

[25] Maria NǍśremark, Susanna Sternberg Lewerin, Linda Ernholm, and Jenny FrǍssling. Swedish Farmers' Opinions about Biosecurity and Their Intention to Make Professionals Use Clean Protective Clothing When Entering the Stable. *Frontiers in veterinary science*, 3, 2016.

[26] Pierre-Alexandre Beloeil, Claire Chauvin, Karine Proux, Christelle Fablet, François Madec, and Ahmadou Alioum. Risk factors for Salmonella seroconversion of fattening pigs in farrow-to-finish herds. *Veterinary Research*, 38(6):835–848, 2007.

[27] Annette Backhans, Marie SjǍslund, Ann Lindberg, and Ulf Emanuelson. Biosecurity level and health management practices in 60 Swedish farrow-to-finish herds. *Acta Veterinaria Scandinavica*, 57(1), 2015.

[28] Manon Racicot, Daniel Venne, AndrǍŠ Durivage, and Jean-Pierre Vaillancourt. Description of 44 biosecurity errors while entering and exiting poultry barns based on video surveillance in Quebec, Canada. *Preventive Veterinary Medicine*, 100(3):193–199, 2011.

[29] Eugene M Caruso, Daniel T Gilbert, and Timothy D Wilson. A Wrinkle in Time: Asymmetric Valuation of Past and Future Events. *Psychological Science*, 19(8):796–801, 2008.

[30] Yaacov Trope and Nira Liberman. Temporal construal. *Psychological review*, 110(3):403, 2003.

[31] Richard Yi, Kirstin M. Gatchalian, and Warren K. Bickel. Discounting of Past Outcomes. *Experimental and Clinical Psychopharmacology*, 14(3):311–317, 2006.

[32] Caroline Millman, Rob Christley, Dan Rigby, Diana Dennis, Sarah J. O'Brien, and Nicola Williams. 'Catch 22': Biosecurity awareness, interpretation and practice amongst poultry catchers. *Preventive Veterinary Medicine*, 141:22 – 32, 2017.

[33] Masanao Toda. The design of a fungus-eater: A model of human behavior in an unsophisticated environment. *Behavioral Science*, 7(2):164–183, 4 1962.

[34] Vernon L. Smith. Experimental Economics: Induced Value Theory. *The American Economic Review*, 66(2):274–279, 1976.

[35] Colin Camerer and Robin Hogarth. The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework. *Journal of Risk and Uncertainty*, 19(1):7–42, 1999.

[36] Lisa Cheong, Susanne Bleisch, Allison Kealy, Kevin Tolhurst, Tom Wilkening, and Matt Duckham. Evaluating the impact of visualization of wildfire hazard upon decision-making under uncertainty. *International Journal of Geographical Information Science*, 30(7):1377–1404, 2016.

[37] Charles A Holt and Susan K Laury. Risk aversion and incentive effects. *American economic review*, 92(5):1644–1655, 2002.

[38] David B Porter. Computer games: Paradigms of opportunity. *Behavior research methods, instruments, & computers*, 27(2):229–234, 1995.

[39] Daniel Boduszek, Agata Debowska, Adele D Jones, Minhua Ma, David Smith, Dominic Willmott, Ena Trotman Jemmott, Hazel Da Breo, and Gillian Kirkman. Prosocial video game as an intimate partner violence prevention tool among youth: A randomised controlled trial. *Computers in Human Behavior*, 93:260–266, 2019.

[40] Erika Montanaro, Lynn E Fiellin, Tamer Fakhouri, Tassos C Kyriakides, and Lindsay R Duncan. Using videogame apps to assess gains in adolescents' substance use knowledge: new opportunities for evaluating intervention exposure and content mastery. *Journal of medical Internet research*, 17(10):e245, 2015.

[41] Scott C. Merrill, Christopher J. Koliba, Susan M. Moegenburg, Asim Zia, Jason Parker, Timothy Sellnow, Serge Wiltshire, Gabriela Bucini, Caitlin Danehy, and Julia M. Smith. Decision-making in livestock biosecurity practices amidst environmental and social uncertainty: Evidence from an experimental game. *PLoS ONE*, 14(4), 2019.

[42] Eric M. Clark, Scott C. Merrill, Luke Trinity, Gabriela Bucini, Nicholas Cheney, Ollin Langle-Chimal, Trisha Shrum, Christopher Koliba, Asim Zia, and Julia M. Smith. Using digital field experiments to elicit risk mitigation behavioral strategies for disease management across agricultural production systems, 2019.

[43] Scott C. Merrill, Susan Moegenburg, Christopher J. Koliba, Asim Zia, Luke Trinity, Eric Clark, Gabriela Bucini, Serge Wiltshire, Timothy Sellnow, Deanna Sellnow, and Julia M. Smith. Willingness to Comply With Biosecurity in Livestock Facilities: Evidence From Experimental Simulations. *Frontiers in Veterinary Science*, 6:156, 2019.

[44] Johanne Ellis-Iversen, Alasdair J.C. Cook, Eamon Watson, Mirjam Nielen, Lesley Larkin, Marion Wooldridge, and Henk Hogeveen. Perceptions, circumstances and motivators that influence implementation of zoonotic control programs on cattle farms. *Preventive Veterinary Medicine*, 93(4):276–285, 2010.

[45] Brian E. Roe. The Risk Attitudes of U.S. Farmers. *Applied Economic Perspectives and Policy*, 37(4):553–574, Aug 2015.

[46] Icek Ajzen. The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2):179–211, 1991.

[47] I.M. Lipkus and J.G. Hollands. The visual communication of risk. *Journal of the National Cancer Institute. Monographs*, 1999(25):149–163, 1999.

[48] Paul Slovic, Melissa L. Finucane, Ellen Peters, and Donald G. Macgregor. Risk as Analysis and Risk as Feelings: Some Thoughts about Affect, Reason, Risk, and Rationality. *Risk Analysis*, 24(2):311–322, 2004.

[49] Paul Slovic, Ellen Peters, Melissa L. Finucane, and Donald G. Macgregor. Affect, Risk, and Decision Making. *Health Psychology*, 24(4S):S35–S40, 2005.

[50] Sabine M. Marx, Elke U. Weber, Benjamin S. Orlove, Anthony Leiserowitz, David H. Krantz, Carla Roncoli, and Jennifer Phillips. Communication and mental processes: Experiential and analytic processing of uncertain climate information. *Global Environmental Change*, 17(1):47–58, 2007.

[51] A Tversky and D Kahneman. Judgment under Uncertainty: Heuristics and Biases. *Science (New York, N.Y.)*, 185(4157), 1974.

[52] Anthony G Patt and Daniel P Schrag. Using specific language to describe risk and probability. *Climatic change*, 61(1-2):17–30, 2003.

[53] Vivianne HM Visschers, Ree M Meertens, Wim WF Passchier, and Nanne NK De Vries. Probability information in risk communication: a review of the research literature. *Risk Analysis: An International Journal*, 29(2):267–287, 2009.

[54] Joachim Scholderer and Nina Veflen. Social norms and risk communication. *Trends in Food Science & Technology*, 84:62–63, 2019.

[55] Daniel Kahneman and Amos Tversky. PROSPECT THEORY: AN ANALYSIS OF DECISION UNDER RISK. *Econometrica (pre-1986)*, 47(2), 1979.

[56] Toshio Yamagishi, Nobuhiro Mifune, Yang Li, Mizuho Shinada, Hirofumi Hashimoto, Yutaka Horita, Arisa Miura, Keigo Inukai, Shigehito Tanida, Toko Kiyonari, et al. Is behavioral pro-sociality game-specific? pro-social preference and expectations of pro-sociality. *Organizational Behavior and Human Decision Processes*, 120(2):260–271, 2013.

[57] E Kristensen and EB Jakobsen. Challenging the myth of the irrational dairy farmer; understanding decision-making related to herd health. *New Zealand Veterinary Journal*, 59(1):1–7, 2011.

[58] Theo Offerman and Andrew Schotter. Imitation and luck: an experimental study on social sampling. *Games and Economic Behavior*, 65(2):461–502, 2009.

[59] Ulf Toelch, Matthew J Bruce, Marius TH Meeus, and Simon Matthew Reader. Social performance cues induce behavioral flexibility in humans. *Frontiers in psychology*, 2:160, 2011.

[60] Gabriele Paolacci, Jesse Chandler, and Panagiotis G Ipeirotis. Running experiments on amazon mechanical turk. *Judgment and Decision making*, 5(5):411–419, 2010.

[61] Michael Buhrmester, Tracy Kwang, and Samuel D Gosling. Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on psychological science*, 6(1):3–5, 2011.

[62] Tony Parisi. *WebGL: up and running.* " O'Reilly Media, Inc.", 2012.

[63] Da Levinthal and Jg March. The myopia of learning. *Strategic Management Journal*, 14:95–112, 1993.

[64] R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2019.

[65] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag, New York, 2016.

[66] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*, 2014.

[67] Kenneth P Burnham and David R Anderson. *Model selection and multimodel inference : a practical information-theoretic approach.* Springer, New York, 2nd ed.. edition, 2002.

[68] Kenneth P Burnham and David R Anderson. Multimodel inference: understanding AIC and BIC in model selection. *Sociological methods & research*, 33(2):261–304, 2004.

[69] Hirotogu Akaike. *Information Theory and an Extension of the Maximum Likelihood Principle.* Springer New York, New York, NY, 1973.

[70] William C Aird. Phenotypic heterogeneity of the endothelium: Ii. representative vascular beds. *Circulation research*, 100(2):174–190, 2007.

[71] William C Aird. Endothelium in health and disease. *Pharmacological reports*, 60(1):139, 2008.

[72] Maryam Nakhaei-Nejad, Maikel Farhan, Anahita Mojiri, Hosna Jabbari, Allan G Murray, and Nadia Jahroudi. Regulation of von willebrand factor gene in endothelial cells that are programmed to pluripotency and differentiated back to endothelial cells. *STEM CELLS*, 37(4):542–554, 2019.

[73] Lina Wang, Weijun Su, Wei Du, Yang Xu, Lijun Wang, Deling Kong, Zhongchao Han, Guoguang Zheng, and Zongjin Li. Gene and microrna profiling of human induced pluripotent stem cell-derived endothelial cells. *Stem Cell Reviews and Reports*, 11(2):219–227, 2015.

[74] Xuechong Hong, Alexandra Le Bras, Andriana Margariti, and Qingbo Xu. Reprogramming towards endothelial cells for vascular regeneration. *Genes & diseases*, 3(3):186–197, 2016.

[75] Doulaye Dembélé and Philippe Kastner. Fold change rank ordering statistics: a new method for detecting differentially expressed genes. *BMC bioinformatics*, 15(1):14, 2014.

[76] Travis E Oliphant. Python for scientific computing. *Computing in Science & Engineering*, 9(3):10–20, 2007.

[77] Shihori Tanabe, Kazuhiko Aoyagi, Hiroshi Yokozaki, and Hiroki Sasaki. Regulation of ctnnb1 signaling in gastric cancer and stem cells. *World journal of gastrointestinal oncology*, 8(8):592, 2016.

[78] Aja Harding, Elizabeth Cortez-Toledo, Nataly L Magner, Julie R Beegle, Dane P Coleal-Bergum, Dake Hao, Aijun Wang, Jan A Nolta, and Ping Zhou. Highly efficient differentiation of endothelial cells from pluripotent stem cells requires the mapk and the pi3k pathways. *Stem Cells*, 35(4):909–919, 2017.