

## Article

# Human Factors in Leveraging Systems Science to Shape Public Policy for Obesity: A Usability Study

Philippe J. Giabbanelli <sup>1,\*</sup>  and Chirag X. Vesuvala <sup>2</sup>
<sup>1</sup> Department of Computer Science & Software Engineering, Miami University, Oxford, OH 45056, USA

<sup>2</sup> Independent Researcher, Vancouver, BC V6B 1Y0, Canada

\* Correspondence: giabbapj@miamioh.edu

**Abstract:** **Background:** despite a broad consensus on their importance, applications of systems thinking in policymaking and practice have been limited. This is partly caused by the longstanding practice of developing systems maps and software in the intention of supporting policymakers, but without knowing their needs and practices. **Objective:** we aim to ensure the effective use of a systems mapping software by policymakers seeking to understand and manage the complex system around obesity, physical, and mental well-being. **Methods:** we performed a usability study with eight policymakers in British Columbia based on a software tool (*ActionableSystems*) that supports interactions with a map of obesity. Our tasks examine different aspects of systems thinking (e.g., unintended consequences, loops) at several levels of mastery and cover common policymaking needs (identification, evaluation, understanding). Video recordings provided quantitative usability metrics (correctness, time to completion) individually and for the group, while pre- and post-usability interviews yielded qualitative data for thematic analysis. **Results:** users knew the many different factors that contribute to mental and physical well-being in obesity; however, most were only familiar with lower-level systems thinking concepts (e.g., interconnectedness) rather than higher-level ones (e.g., feedback loops). Most struggles happened at the lowest level of the mastery taxonomy, and predominantly on network representation. Although participants completed tasks on loops and multiple pathways mostly correctly, this was at the detriment of spending significant time on these aspects. Results did not depend on the participant, as their experiences with the software were similar. The thematic analysis revealed that policymakers did not have a typical workflow and did not use any special software or tools in their policy work; hence, the integration of a new tool would heavily depend on individual practices. **Conclusions:** there is an important discrepancy between what constitutes systems thinking to policymakers and what parts of systems thinking are supported by software. Tools may be more successfully integrated when they include tutorials (e.g., case studies), facilitate access to evidence, and can be linked to a policymaker's portfolio.

**Keywords:** chronic disease; conceptual map; policymaking; systems thinking; usability



**Citation:** Giabbanelli, P.J.; Vesuvala, C.X. Human Factors in Leveraging Systems Science to Shape Public Policy for Obesity: A Usability Study. *Information* **2023**, *14*, 196. <https://doi.org/10.3390/info14030196>

Academic Editor: Valentina Casola

Received: 2 February 2023

Revised: 10 March 2023

Accepted: 17 March 2023

Published: 20 March 2023



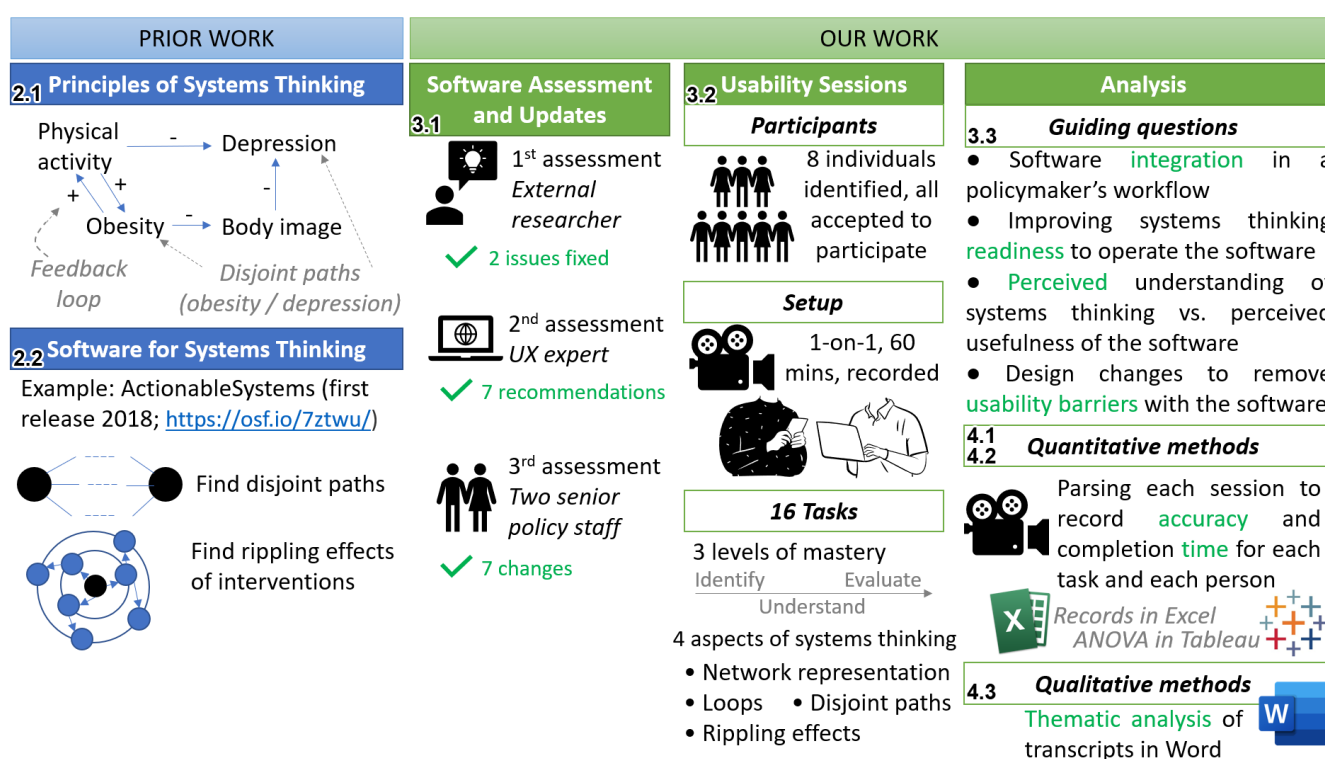
**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Grand societal challenges as varied as suicide prevention [1,2], climate change adaptation [3], or obesity [4,5] are known as *wicked problems*. Since they are not caused by a handful of root causes, they are not easily addressed solely via reductionist thinking strategies that break down an intervention program into a few independent components [6,7]. Rather, wicked problems are the product of *systems*, consisting of interconnected elements whose interactions exhibit non-linear dynamics. A *systems thinking* approach encompasses reductionist strategies and incorporates advances in system dynamics or complexity theory to achieve a holistic view of the “interrelationships between parts and their relationships to a functioning whole” [8,9]. Scholars have emphasized that systems thinking is needed now more than ever [10] to holistically examine complex systems [7]. In particular, they

repeatedly ask “for new tools and approaches to analyze wicked problems and grand challenges” [10].

However, applying systems thinking through the creation of tools such as *systems maps* and accompanying software [11] can be challenging. For example, an information overload [7] can occur when we attempt to comprehend the breadth of a topic such as obesity, which involves domains such as nutrition, psychology, the built environment, or physiology. In addition, key notions of systems thinking such as feedback loops (Figure 1) can be difficult to understand and communicate, especially for audiences that are used to reductionist strategies [7]. Several *mapping software* allow policymakers and intervention designers to collaboratively create a map [12], but fewer provide support the navigation of maps by offering features that seek to address information overload (e.g., decomposition of a system into high-level components that can be *expanded on demand*) and facilitating typical queries for reductionist strategies (e.g., listing concepts involved between an intervention *target* and its evaluation endpoint) [13].



**Figure 1.** Key components of this study along with corresponding subsections.

In a field such as obesity, the historical approach often boiled down to ‘if you build it, they will come’: models and software were released with the intention of supporting policymaking, but policymakers were rarely part of the project and ultimately the uptake of the tools was limited [14]. Specific reasons for a limited impact include a disconnect between how developers *perceived* policymaking needs compared to the practices of policymakers. Indeed, “the capacity of policymakers to use effectively the full range of existing and emerging analytical tools may be limited by their general knowledge and associated understanding and skills” [15]. Consequently, a subtle yet essential paradigm shift occurred in the development of systems science models and software intended to support policymaking efforts. Indeed, studies are gradually stressing the importance of building ongoing partnerships with policymakers and other stakeholders by investing time to *understand their needs and resources* [16,17]. For instance, a study on policymakers’ perceptions emphasized that there should be a visual display [18], but developers cannot simply use visuals that *they* understand, as that may not translate to the needs of end users.

In the case of systems thinking for public health, there has been a disconnect between the practices of model developers and the needs of end users for tangible assistance [19]. On one side, modelers have invested considerable energy in thorough inventories of the factors and relationships that create a complex system. For instance, maps of obesity have covered 98 factors [20], 108 factors [21], and up to 114 factors [22]; each of these maps connected factors through more than a hundred relationships. The *intention* was to provide policymakers with comprehensive tools that support systems thinking approaches. Little was known about how policymakers engaged with systems thinking [19]; hence, assumptions were made and occasionally became problematic. Specifically, the laudable aim that policymakers should be provided with very comprehensive maps of causes and interrelationships implicitly assumed that they wanted such maps and would be able to use them effectively via existing software. However, on the other side, target users viewed the maps as an “almost incomprehensible web of interconnectedness [in which] the scale and number of interactions make it difficult to see how one might use it in any practical way to develop systemic approaches” [23]. For example, a commentary suggested that the Foresight Obesity Map [21] “looks more like a spilled plate of spaghetti than anything of use to policymakers” [24] and this ‘spaghetti map’ analogy has been used for over ten years [25,26]. Said otherwise, a tool intended to give a sense of direction [27] became used as a herald of complexity; hence, portraying obesity as an intractable problem.

As a result of this disconnect, scholars have noted that applications of systems thinking in policymaking and practice have been scarce [19,28], despite a broad consensus on its importance [28] and its increasing presence as a discussion topic in public health [29]. To address this gap, our study evaluates the *usability* of a software that seeks to support policymakers in using a systems map of obesity. Usability is indeed essential to ensure the effective uptake of a tool; however, it has rarely been studied for maps, and only in the context of educational technologies for knowledge assessment [30–32] where software difficulties are occasionally reported [33]. Our study thus contributes to the literature by *evaluating the ability of eight policymakers to interact with systems maps* through recorded sessions that follow best practices in usability testing. We focus on an updated version of the software tool *ActionableSystems* [13], which we developed and released as an open-source package to support systems thinking and policymaking.

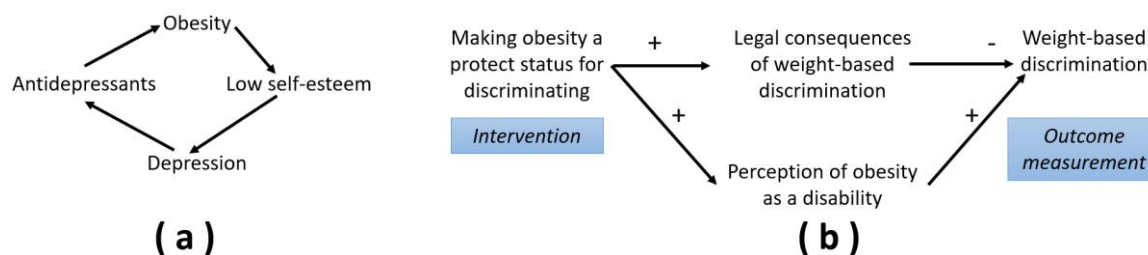
The key parts of our study are organized as shown in Figure 1. Section 2 provides a succinct introduction to the core themes articulated in this paper; these themes include systems thinking and its application to obesity, the previously released *ActionableSystems* software (e.g., its guiding idea and functionalities), and usability testing. These overviews are provided to keep the paper self-contained and can be skipped by readers familiar with some of these notions. Then, in Section 3, we detail our study design, including the design of each usability session and the mixed methods analysis. Our results in Section 4 report key usability metrics (correctness, time to completion), both at the level of individual users and overall, for three common policymaking tasks (identification, evaluation, understanding). A complementary thematic analysis provides in-depth feedback from the users. Our discussion in Section 5 summarizes some of our limitations and offers suggestions for future work.

## 2. Background

### 2.1. Principles and Challenges of Systems Thinking for Obesity Research

Overweight and obesity have historically been oversimplified either to a mere matter of diet and exercise [34,35] or by considering weight-related factors in isolation. However, systems elements such as feedback loops (Figure 2a) have stronger associations with weight categories for individuals than isolated factors [36]. The Foresight Obesity Map and more recent systems maps [20–22,37] have contributed to cementing the notion that overweight and obesity are parts of a complex system involving many other important components, such as physical or mental well-being. In its simplest form, systems thinking starts as the

recognition that overweight and obesity are part of such a system of interacting factors, and that (while convenient) they should not be artificially isolated from this larger system.



**Figure 2.** Common tasks in systems thinking with respect to maps include finding structures such as loops (a) and disjoint paths (b). For policymaking, loops can amplify effects or create inertia, while disjoint paths between an intervention and its measurement can include unintended side effects. The software studied here addresses both structures.

While early strategies for obesity prevention targeted individual behaviors, programs have become increasingly *complexity-oriented* over time [38]. The ongoing search for efficient cross-sectoral public policies regarding healthy weights and healthy living exemplifies different aspects of this complexity [39,40] such as heterogeneity [36,41], interventions at multiple scales [42], or the interaction between individuals and their environments [43]. Although systems approaches are not consistently defined in the literature, they either involve modeling individuals (e.g., via Agent-Based Models [44]) or modeling mechanisms through maps [11], ranging from causal maps and causal loop diagrams [45,46] to more quantitative approaches such as system dynamics [47,48] or fuzzy cognitive maps [41]. Our paper focuses on the ability of policymakers to interact with such maps.

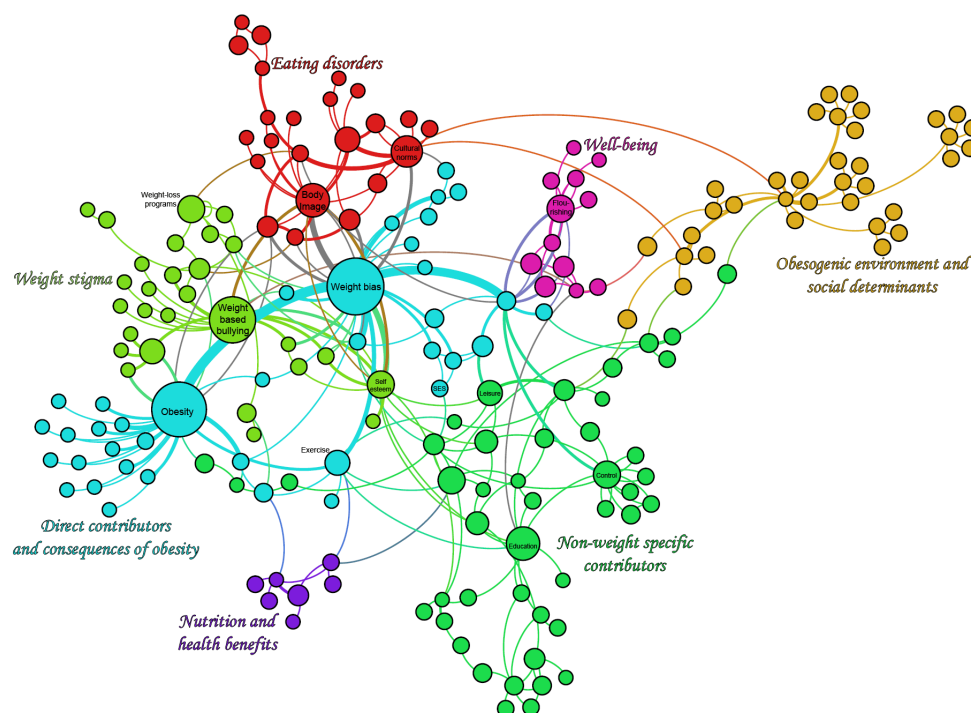
While we tend to think in straight lines of cause and effect relationships, systems science emphasizes that we most often face *feedback loops*: the factor on which we act will trigger changes that will ultimately affect it in the future (Figure 2a). In other words, the variation in a factor propagates through other factors and returns to the initial one to produce another change. Abundant examples of feedback loops in the metabolic aspects of obesity can be found in *Thinking in Circles About Obesity* [49]. However, it is challenging for humans to naturally think of loops [50]; studies found that this challenge did not stem from a lack of expertise (participants were policy specialists), voluntary simplification after deliberation (some models were immediately discussed after presenting a new problem), or a focus on short-range effects (one participant was planning a constitution). Rather, a cognitive explanation was most likely, as confirmed by later findings that people “tend to (un)consciously reduce complexity in order to prevent information overload and to reduce mental effort” [51], which leads to eliminating loops and ignoring multiple paths between two concepts (Figure 2b). These problems will be found even when participants receive extensive training [51,52].

Systems thinking encompasses the *identification* of structures such as loops and paths. It is also a conscious effort to *leverage* such structures to carefully craft interventions on the system and evaluate their effects (e.g., unintended consequences, blocking points). Donella Meadows famously introduced a hierarchy of leverage points in a system [53], which illustrates that thinking of individual factors would be the lowest level of systems thinking, while examining loops is at an intermediate level, and changing goals or paradigms sits at the highest level. The Intervention Level Framework (ILF) was adapted from Meadows’ work with the purpose of bringing systems thinking to policies around overweight and obesity [54]. An application highlighted that policymakers do acknowledge obesity as a complex problem; but at the same time, their strategies are often at a low level of systems thinking [55].

## 2.2. Embodying the Principles of Systems Thinking for Obesity via Software Solutions

The Provincial Health Services Authority (PHSA) of British Columbia reviewed the evidence regarding the relationships between overweight, obesity, and mental well-being [56]. It suggested a paradigm shift by tackling common determinants of weight and health (e.g., weight bias, stigma) to promote the flourishing in mind and body for all (Figure 3). This report stood in contrast with others at the time, as it had the largest number of recommendations at a high level of systems thinking, and the lowest number of recommendations denoting little systems thinking [55]. The report called for additional evidence to connect physical and mental well-being, thus resulting in the creation of a causal map [20]. The design of the map was grounded in a systems thinking perspective, and focused on teasing out the pathways and multiple loops at work in obesity and well-being. Given the interest generated by the map, our team designed and implemented the new open-source *ActionableSystems* software (<https://osf.io/7zwtwu/>; last accessed on 18 March 2023) to facilitate interacting with the map [13]. The present manuscript focuses on improving and evaluating the usability of this software. We previously presented the first version of the software through a dedicated workshop at the 4th Canadian Obesity Summit [57]. The first version included four tools, each aiming at supporting one of the hallmarks of systems thinking:

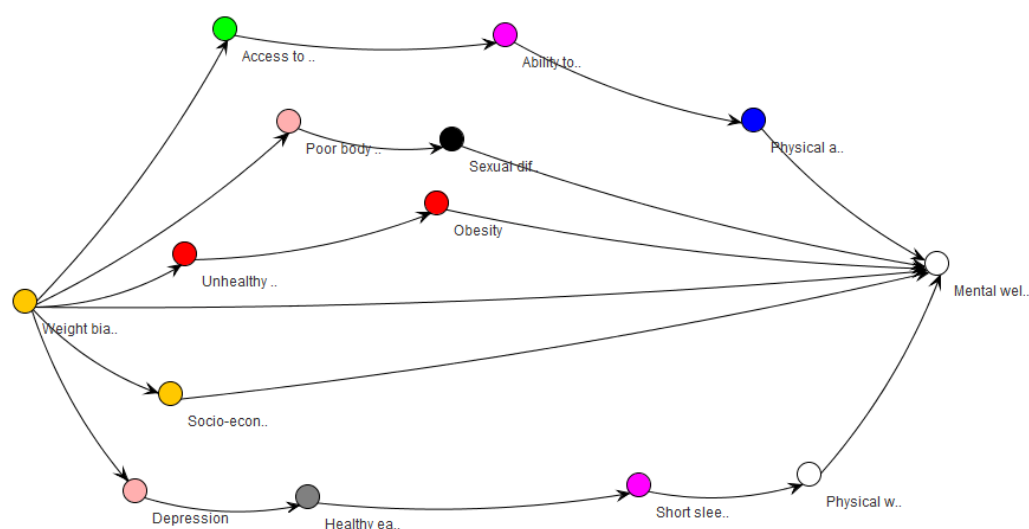
- An *exploration tool* (Figure 4) to interactively visualize the factors and their relationships. For example, one may ask “my work is on weight stigma, how does it relate to other parts? What increases stigma?”;
- A tool for *rippling consequences* (Figure 5) to identify what would be (in)directly affected by a policy. For example, one may ask “we aim at creating vibrant communities, what will that directly impact?”;
- A *pathway tool* (Figure 6) to find the multiple pathways connecting two factors. For example, one may ask “I promote a policy against bullying, and I am interested in the outcome on well-being. What is between them?”;
- An *importance tool*, which ranks factors by how central they are in the system. For example, one may look for what lies at the core of the whole system, which may lead to questioning paradigms about the system itself.



**Figure 3.** Visualization of the PHSA report [56] as a causal map, where each factor is shown as a node (circle). This map was extended into a version with 98 factors [20]. Sizes indicate centrality and colors







**Figure 6.** Policymakers are often interested in measuring the consequences of an intervention. In the first version of *ActionableSystems*, this is supported by choosing an intervention node (e.g., weight bias) and a measurement target (e.g., mental well-being). The intervention will be placed on the left, and all paths that lead to the target can be followed left-to-right in order to identify all parts of the system that would be impacted by the intervention. This visual can reveal unintended consequences, such as impacted sleep duration. Factors that belong to different themes are shown in different colors (e.g., psychological constructs are light pink while well-being constructs are white).

A brief overview video, available at <https://www.youtube.com/watch?v=OdKJW8tNDcM> (last accessed on 18 March 2023), summarizes the rationale for the first version of the software and shows its ability to generate interactive visualizations in response to practical policy questions.

### 2.3. Purposes and Core Concepts of Usability Testing

Some software work well and are easy to use, while others have *bugs* or are poorly *designed* such that they are not as intuitive as one might like. Traditional software testing is employed to find bugs in the code. Usability testing is employed to find a design issue that, while technically correct, makes the software less user-friendly and harder to use. It ensures that the design of the software allows users to get the most out of the tool.

Usability testing is generally done by observing a representative user and asking them to use the software, in contrast to automated testing procedures for bugs [58]. The user might run into problems, such as not knowing where to click or what the different buttons mean, or they may dislike aspects such as the color of the background. After multiple sessions (each with a different user) are conducted, the usability specialist will look for patterns in the feedback; that is, issues that occurred in multiple sessions. These are the usability issues that need to be addressed. Typically, one does not need to fix/address usability issues that only occurred in a single session, unless it is deemed to be a serious issue by the project team. Usability “testing” can be misunderstood as a final step of product development. However, usability evaluations are typically done as part of the design stage within a product development *cycle* [59]. In other words, the software design and its usability evaluation occur multiple times.

Some of the questions addressed by performing a usability test include: does the software meet the users’ *needs*? Does it behave as the user *expects*? What tasks are *frustrating* for the user to perform? How can we make the software more *useful* for the user? These questions also offer insight for higher-level inquiries, such as whether the users’ mental model is aligned with the conceptual model embedded in the software. In other words, usability testing checks for discrepancies between the users’ ideas of how the software should work versus how the developer wants the user to use the software.

There are five key concepts in usability testing. *Findability* is the idea that the user will be able to find an option or a control. That is, if a feature is not immediately visible, is it where the user expects to find it? For example, the save feature is usually under the 'File' menu hence a user may look there first. The *mapping of controls to functions* needs to be understood such that users can predict the outcome of an action before performing it. For instance, changing the font from a smaller number to a larger number makes the font larger. The size of the font and the number are linked in the software and in the users' minds. *Feedback* is helpful for users. There should be some visual and/or auditory confirmation that the action has been registered by the computer and that it is being carried out such that progress is being made. Think of the loading bar when installing a program: if no feedback is given, then the user may perform the action again without realizing what is happening; saving a file multiple times may be fine but hitting the delete key repeatedly may remove more files than intended. Some aspects of user behavior may be *modelled*. For example, how long it takes to move the mouse from one part of the screen to a target on the other side of the screen is a function of the distance and the size of the target. Other aspects of software can be evaluated by doing an *expert walk-through* [60], which can identify general issues (e.g., font too small, unusual icons) but does not replace testing with an actual user.

#### 2.4. Conducting Usability Testing

As a scientific study, usability requires a *test plan* before any sessions are run. The test plan should cover all the details of the study, including participant profile, link to recruit participants, how participants will be rewarded, research questions that are being investigated, specific tasks and questions that will be asked of the participants, what equipment will be used, and the schedule.

Once the interface has been designed, the next step is to find participants and bring them in to do tasks on the computer while being observed by the researcher. Finding participants can be challenging, as numerous persons may have to be contacted to find those who fit the description and are available, they may later cancel or simply not show up, and they must be representative of the target user population. Second, participants should be acknowledged, particularly if they are internal participants; otherwise, one runs the risk of the study being exploitative [61]. In other words, participants are giving us valuable feedback and their contributions and efforts should not be taken for granted. Preparing a usability study also involves logistics to bring in participants (e.g., accessibility, parking).

Once a participant is in the room, the study can start. It is important to keep in mind that the study tests the software *through* the participant: it is not testing the participants themselves. The (concurrent) think aloud protocol along with screen and voice recording software are one approach to gather feedback by encouraging participants to verbalize their thoughts when performing tasks [59,62]. Both the consent forms and the facilitator should make it clear to participants that they can stop the study at any time or skip a question or pause if they need to. This standard procedure considers participants as guests whose participation is optional rather than being 'hostages' of the study. It also seeks to reduce the stress of being watched while performing tasks. Since usability studies end at the allotted time (even if there are some incomplete tasks left) and participants can skip questions, the researcher does *not expect all tasks to be completed by every participant*.

The specific tasks to test can each be written on a separate piece of paper. Participants are asked to read each question out loud and reminded to also think out loud while performing it. The "one task per sheet" setup helps the participant stay calm and not get overwhelmed with all the questions at once. Indeed, the study design strives for authenticity: it seeks to limit stress or the feeling of being pressed for time, particularly because time is often tested for. Although 'tasks' are part of the research jargon, it can have a negative association in the users' minds; thus, the word is avoided during the session. As is common practice in qualitative research, questions should not be phrased as leading questions. For example, "tell me what you liked about the software" would be rephrased as "what did you think about the software?"



Tasks can range from very broad (“how would you solve obesity”) to very directed (“click on the factor Stress, then follow the connection to eating”). The tasks depend on what needs to be investigated. A specific task serves to see how a user interacts with a specific part of the program, in contrast to a general task for which we do not know which features can be used. A typical first task is “tell me what you think this software is about?”, after letting the user view the screen for about 5 s.

Participants may ask questions during or after the session, and these could be met with different answers from the researcher. Indeed, the *researcher acts as a facilitator* rather than an authority during the session. Thus, questions during the session may be answered by “I will answer that at the end of the session” to *avoid giving away what the test seeks to establish*. In other words, the researcher needs to remain neutral during the session by being humble, respectful, observant, and patient.

After the sessions are completed, the *analysis* takes place to find *commonalities of issues* among users and to *prioritize* which ones to address next. The analysis starts by re-watching all video records of the user sessions. Each issue, or unexpected behavior of the software, is noted along with recommendations that the users may have offered. Since users are articulating what they were thinking, this can also be analyzed through qualitative methods. Issues that arose across multiple sessions are the ones the team must address. Findings are then discussed within the larger software development team through brainstorming sessions. It serves to make the team more user centric. Screenshots of where the users were having trouble and specific quotes from users (both positive and negative) are often used: they accurately depict what to solve and provide evidence about why it needs to be solved. Changes made to the software are then summarized.

### 3. Methods

#### 3.1. Software Assessment and Updates

The software was assessed and updated three times before the version that served for the usability study. The **first assessment** was conducted by a research staff who had not been previously involved with the project. It was observed that edges were colored in green and red with a symbol (+ or −), indicating that the causal link created either an increase or a decrease. Red was the color associated with an increase (+) while green was associated with decrease (−). A recommendation was made to swap the colors (“red and green line convention”), which was implemented. A bug (i.e., software halting) was found when attempting to find pathways between two factors and mistakenly using the same for both ends. This was addressed by allowing the user to pick any factor as a starting point, but removing the picked factor from the list of valid endpoints.

The **second assessment** was performed by a usability expert who was external to our software development. Four issues were identified:

- a. The conceptual model was not conveyed by the interface. The initial screen, when the software is launched, was overwhelming. Users were not given a chance to mentally process what was going on. The effect of such clutter leads very quickly to mental fatigue. For a user who has other tasks to get to, this can quickly lead to “I’ll figure this out later” or at worst “This is too complicated”.
- b. Unexpected techniques were needed to navigate the map with a mouse. It is common to click and drag on an interface in order to move to a new area of focus. Given that this map is large, and users will need to move around and expand different areas, they will initially expect that clicking and dragging will allow them to drag the canvas (map). This interaction style, however, was not available. Additionally, it was noticed that scrolling the mouse’s wheel upward zooms out the view, which is the opposite of what is commonly found in other software (e.g., Google Maps).
- c. The search function displayed more than was needed. Due to the interface being very busy, users will rely heavily on the “Search” feature. While the search feature does expand the factor of interest, it still displays the entire map on the screen. This will, again, reinforce a sense that the software is “unnecessarily complex”.

- d. Several bugs hindered usability. Instructions mention a right click which does not seem to be enabled.

Seven recommendations were formulated and implemented, when possible, to address these issues. Recommendations are reported with respect to the aforementioned issues (e.g., a-3 is the third recommendation for issue a):

- a-1 Rather than show all relationships at the beginning, we displayed a simple relationship between a few factors only (with other factors closed; therefore, not overwhelming the user). We accomplished this by starting the map at the theme level, with no theme expanded; thus, minimizing the relationships displayed.
- a-2 We included a persistent legend that explains the different shapes, thickness of lines, and arrowhead directions.
- a-3 We created an initial “welcome screen” that briefly explains what this software is to be used for/what value the software hopes to deliver. The popup included an option for “Do not show this again”.
- b-1 We implemented a click and drag interaction style for the map.
- b-2 It would be preferable to allow users to zoom in by scrolling the mouse wheel upward and zoom out by scrolling downward. However, this recommendation could not be implemented, as this behavior was controlled by a third-party library.
- c-1 When search is used, it is desirable to display a “zoomed-in” view of the resulting factor; therefore, hiding most of the map. However, the third-party library does not allow this functionality.
- d-1 We enabled the right click.

The third assessment was done by two senior PHSA staff together. Updates resulting from this assessment are listed in Table 1; they cover the software as well as the map.

**Table 1.** Software changes made after a review by senior staff at the Provincial Health Services Authority (PHSA) of British Columbia, Canada.

Before Assessment by Senior PHSA Staff	After Assessment
The software starts in a small window, creating a very packed map with many labels overlapping.	The software starts in full screen and the map occupies all available space, thus decreasing cluttering.
Some labels on the legend (e.g., Food production) differ from those visible on the map (e.g., Production).	Labels on the map are coherent with those in the legend.
Only the exploratory tool has a detailed legend.	All tools have a detailed legend.
There is a tool to find the most important factor, resulting in a table of numbers. The tool requires extensive expertise to be used but draws in user attention because of the idea of ‘importance’.	The tool was deactivated.
Only the exploratory tool uses the red and green line convention for connections.	All tools use the red and green line convention.
The ‘find the impact tool’ does not run with its default values: users must select one.	The ‘find the impact tool’ runs even if users do not change the default value.
‘Quality of sport infrastructure’ was duplicated in the map.	‘Quality of sport infrastructure’ appears only once.

### 3.2. Session Design

The goal of this study is to ensure that our improved *ActionableSystems* software can effectively be used to support policymakers in understanding and managing the complex system around obesity, physical, and mental well-being. This overarching goal can further be divided into two specific aims: (1) understanding what policymakers need in order to navigate complex systems and (2) assessing and improving the software usability. The design of the sessions seeks to inform these two aims. Each session was designed to last 60 min, based on best practices for usability testing (see the Background). We

identified eight individuals who qualify as policymakers and agreed to participate in the study (Appendix A). Each session was introduced by following a script (Appendix B), then three consecutive activities were performed: an initial interview (to understand the participant's role and current systems thinking), the usability study itself involving using the software to perform tasks (to assess software usability), and overall thoughts (to brainstorm suggestions such as alternative audiences for the software). The script for all three activities is provided in Appendix C. Our analysis accounts for the different format (e.g., semi-structured interview vs usability assessment) of the tasks, since the methodology to analyze a semi-structured interview differs from the one employed in a usability assessment.

The usability study received particular attention as it is the cornerstone of this work. The tasks were chosen with respect to two dimensions (Table 2): the aspect of systems thinking that was involved (e.g., unintended consequences, multiple pathways) and the level of mastery of the task. The difficulty of a task was sorted using Bloom's Taxonomy [63], which is very commonly used to classify learning objectives into levels of mastery and has been applied to maps [64–66].

**Table 2.** Matrix of questions by level of mastery (using Bloom's taxonomy) and systems thinking aspect.

	Identification	Understanding	Evaluation
Network Representation	<ul style="list-style-type: none"> <li>Find the "Physiology" theme.</li> <li>Find factor "emotional eating".</li> <li>How many factors directly influence "emotional eating"?</li> <li>What directly influences "emotional eating"?</li> <li>How many factors does "emotional eating" influence?</li> <li>What does "emotional eating" directly influence?</li> </ul>	<ul style="list-style-type: none"> <li>Does "depression" influence "emotional eating"?</li> </ul>	
Loops	<ul style="list-style-type: none"> <li>How many factors are in the path starting at "Insulin resistance", going through "Cells intake of fatty acids" and a few others, ended at "Insulin resistance" again?</li> </ul>	<ul style="list-style-type: none"> <li>Please describe the relationship between "healthy eating" and "short sleep duration"?</li> </ul>	<ul style="list-style-type: none"> <li>Is this a positive feedback loop or a negative feedback loop?</li> </ul>
Multiple Pathways	<ul style="list-style-type: none"> <li>Does "Weight Bias" indirectly influence "Blood pressure"?</li> </ul>	<ul style="list-style-type: none"> <li>How many pathways are there from "Healthy eating" to "Blood pressure"?</li> </ul>	<ul style="list-style-type: none"> <li>How does the relationship between "Exercise" and "Blood pressure" compare to the relationship between "Appetite" and "Blood pressure"?</li> </ul>
Secondary and Tertiary Effects	<ul style="list-style-type: none"> <li>What else might you think about while considering a policy on "Public health messaging around thinness"?</li> </ul>		<ul style="list-style-type: none"> <li>What might be an unintended consequence of changing "Medications"?</li> <li>You are considering impacting "Late night TV watching", what else needs to be considered for it to have a maximum effect on "Healthy eating"?</li> </ul>

### 3.3. Analysis Plan

To address our two specific aims, our analysis includes a set of four guiding questions. As our analysis takes an iterative approach, the questions can be partly inter-related such that findings regarding a question can prompt further investigation into another. This iterative research process is typical when performing qualitative studies.

First, *how could the software be best integrated into a policymaker's workflow?* Contributing elements will be the details of the interviewers' current and ideal workflows (from the starting interview), suggestions as to who might benefit from the software (from the closing interview), and performance on using the software for specific goals (from the usability session).

Second, *how can we bridge the level of readiness from our audience in systems thinking with what is required to operate the software?* The software provides solutions to navigate and manage a causal map or find multiple pathways between factors related to obesity and well-being. This assumes a high level of readiness with systems thinking and abstract representations as maps. In order for the software to be most useful, it is necessary to understand where our target audience is at, and what needs to be done to address any identified gap. Contributing elements will be the interviewees' conceptualizations of systems (from the starting interview) and an examination of unfamiliar notions when using the software (from the usability session).

Third, *what is the relationship between the perceived understanding of systems thinking and the perceived usefulness of the tool?* Both the starting and closing interviews will be examined to find whether there is a relationship between an interviewees' levels of readiness and his/her recommendation for who might benefit from the tool or what tasks it would be best for. This will inform the next iteration of the software design, and particularly the content/positioning of tutorials.

Finally, *what design changes need to take place to remove any usability barrier in using the tool?* While the second question focused on the interviewees' levels of readiness for the specific concepts involved in the tool, this question will focus on current design choices (e.g., look/location of button, content of legend) that created obstacles when using the software. This will draw on the usability sessions.

Our analysis is performed using mixed methods. Quantitative performance indicators (e.g., the accuracy or time of each user on each task) inform the first and last questions (software integration and usability barriers). The qualitative, transcribed interviews inform all four questions, for instance through a thematic analysis. Our quantitative data is stored in Excel spreadsheets and analyzed either via Excel (e.g., for ANOVA) or Tableau (e.g., for correlations). Transcribed interviews are stored and analyzed in Word.

When performing the analyses, we did not have access to additional variables, such as a policymaker's demographic information (e.g., age), years of experience, or primary area of expertise. As shown in our results (next section), the stability of our results across participants suggests that such additional variables would not play roles as mediating factors.

## 4. Results

### 4.1. Overview

Quantitative measures from the usability session are provided both at the level of *individual* participants (Figures 7 and 8; Table A2 in Appendix D) and *across* all participants (Figures 9 and 10). The accuracy shows whether a participant provided the right answer (Figure 7—green) or not (Figure 7—orange); this applies only to tasks that had a list of correct answers (e.g., 'does depression influence emotional eating', 'how many pathways are there') but not to open-ended questions. The aggregate summary (Figure 9) shows that questions on loops and multiple pathways are answered mostly correctly but struggles are found in network representation. While loops and multiple pathways lead to more correct answers, participants also chose to spend significantly more time on these questions (Figures 8 and 10). Additional statistical analyses of the usability session are provided in the next subsection.

## Accuracy by User



Figure 7. Correct (green) or incorrect (orange) answers for each user and task.

## Time to Complete by User

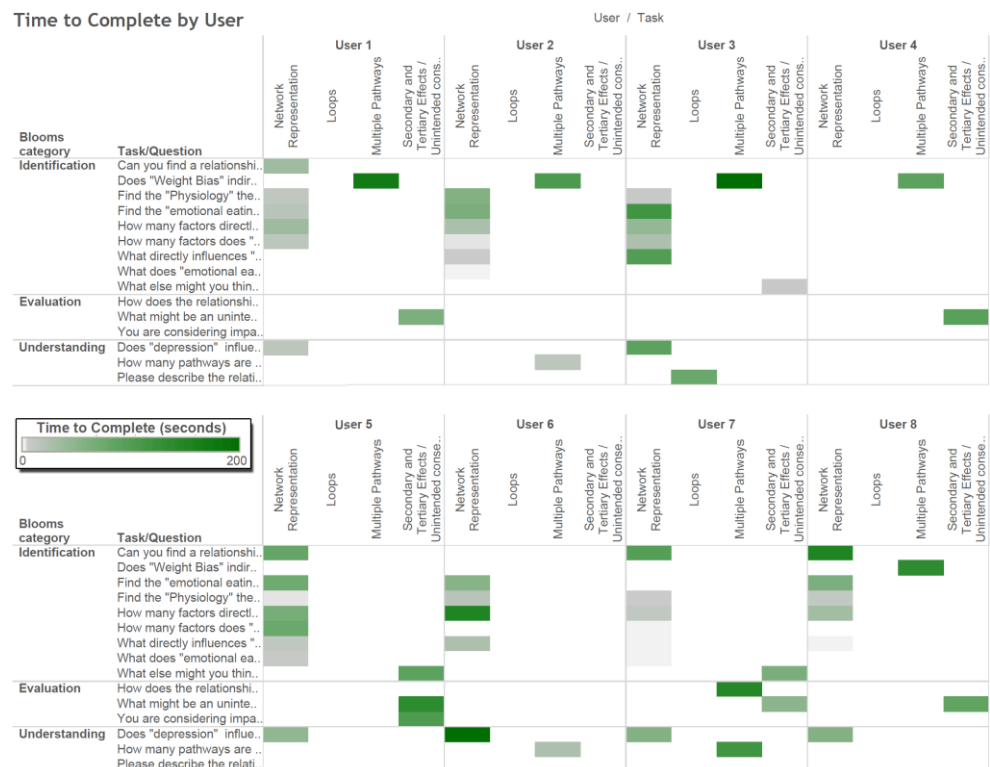
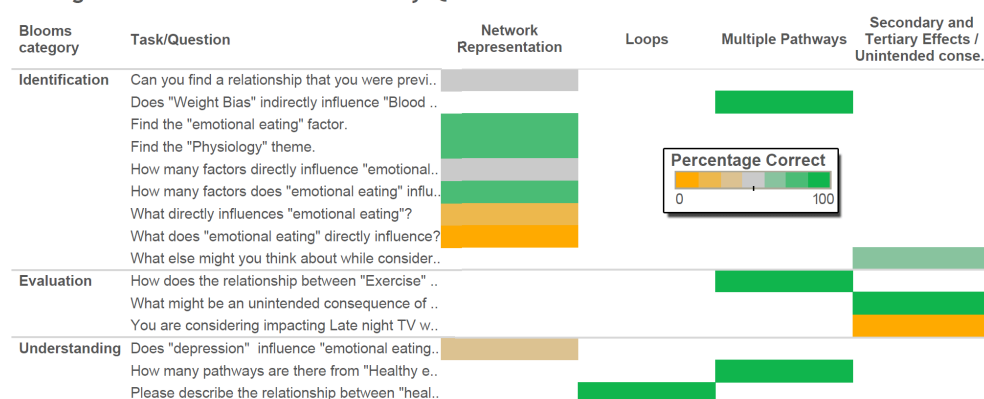


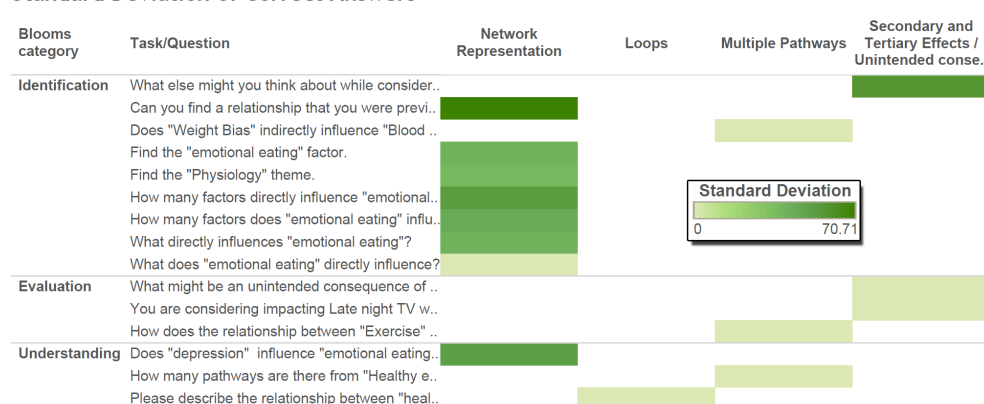
Figure 8. Time to complete each task for each user. Darker colors indicate a longer time to completion.



## Average Number of Correct Answers by Question



## Standard Deviation of Correct Answers



**Figure 9.** Average percentage of correct answers (top) and standard deviation (bottom) across all users.

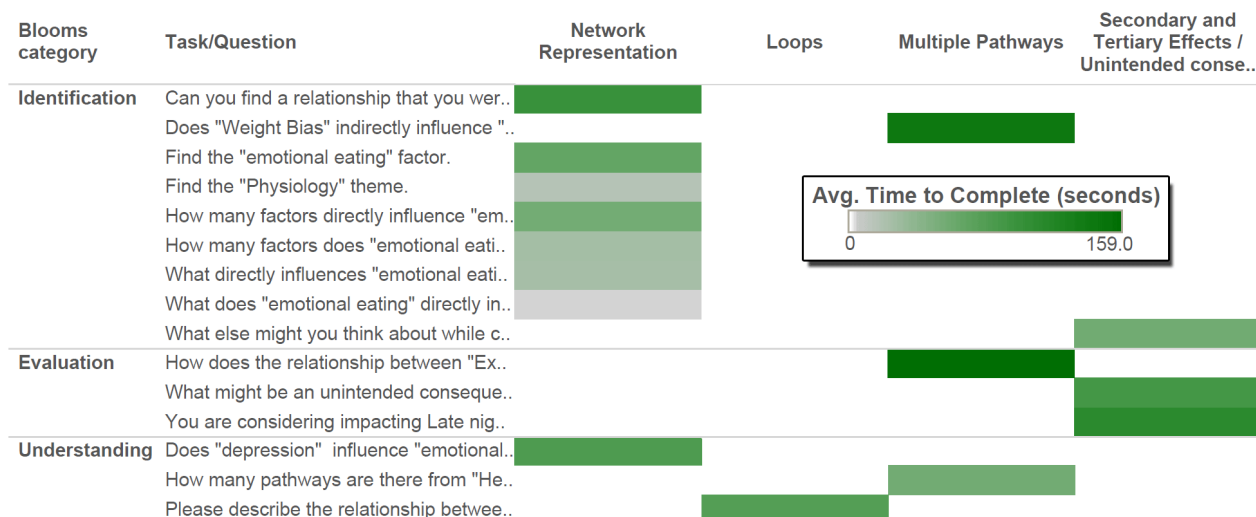
Based on the usability sessions, we formulated a total of 32 recommendations, (Appendix E). These were overwhelmingly at the lowest level of Bloom's mastery taxonomy (identification,  $n = 26$ ) and were mostly in terms of network representation ( $n = 26$ ). These recommendations were driven by the problems that our participants experienced, which were found both via quantitative measures (i.e., time to completion of a task, accuracy in completing it) and qualitative measures (e.g., participants' verbal feedbacks when completing the task, participants visual journeys on the screen).

The thematic analysis of the systems thinking interview (administered before the usability session) revealed five themes, examined in the last subsection; quotes are provided in Appendix F to illustrate the feedback from participants. We found that participants were familiar with systems thinking but may use the word to represent lower level concepts, such as interconnectedness, rather than higher level ones (e.g., feedback loops). Consequently, the fact that no recommendation was made about feedback loop should not be interpreted as meaning that none can be made; rather, that the participants were not at a sufficient level of mastery to face that concept. Similarly, the fact that recommendations overwhelmingly concern the identification and network representation tasks suggests that participants were experiencing general issues in the use of causal maps.

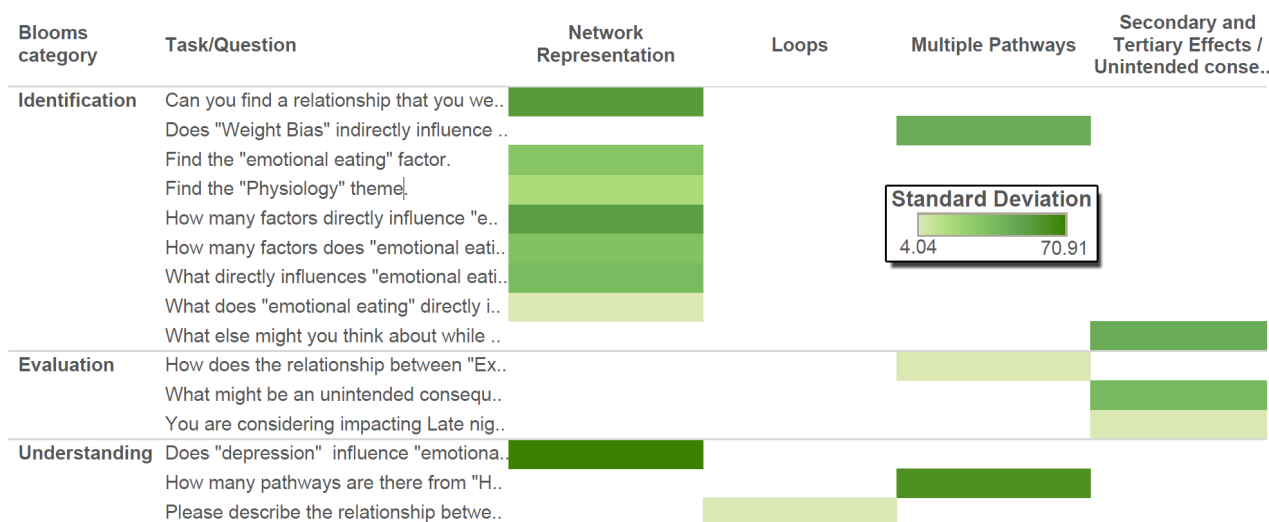
After performing the usability study, we elicited additional thoughts on the software from the participants, leading to six themes. Participants reported that the software could have value in exploring and planning; however, they also highlighted the need to provide practical case studies to avoid having the software appear as a theoretical study of complex systems. Among users who would benefit from the software in the future, participants recommended health authorities and staff in a leadership position. While dieticians were also recommended, the dieticians themselves mentioned different people. This suggests

that the key target audience would be exploring and managing complex systems, rather than being an expert in a specific area of the system (e.g., nutrition).

### Time to Complete, Average



### Time to Complete, Standard Deviation



**Figure 10.** Average time to completion (**top**) and standard deviation (**bottom**) across all users.

#### 4.2. Statistical Analyses

Further understanding of the results of the usability study raised two questions. First, do the results depend on the participant, or are participants having a very similar experience? We analyzed whether there was a statistically significant difference between the participants, either in terms of response time or of accuracy. The procedure (detailed below) shows that there is no statistically significant difference. Second, if users were able to better interact with the network representation, would it benefit performance on other tasks? We analyzed whether the speed or accuracy on network representation tasks was indicative of performance on other tasks in terms of either speed or accuracy. We could not analyze the impact on accuracy in other tasks because only network representation tasks had sufficient variation in the number of incorrectly answered questions. Our procedure (detailed below) shows that there is no connection between network representation tasks and time spent on doing other tasks.

Specifically, we performed a one-sided ANOVA to assess whether there was a statistical difference between the means of the users' times (Table 3). Since  $F < F_{crit}$ , we fail to reject the null hypothesis and we conclude that the means are the same. That is, our users are comparable in terms of time spent on tasks. We also performed a one-sided ANOVA to assess whether there was a statistical difference between the means of the users' accuracies where correct answers were coded as 1 and incorrect answers were coded as 0 (Table 4). Again,  $F < F_{crit}$ ; hence, our users are comparable in terms of accuracy. Note that, to *anonymize the performances* of specific users, the number assigned to a user may not be the same across tables or figures.

**Table 3.** Results from the ANOVA on users' times with  $\alpha = 0.1$ .

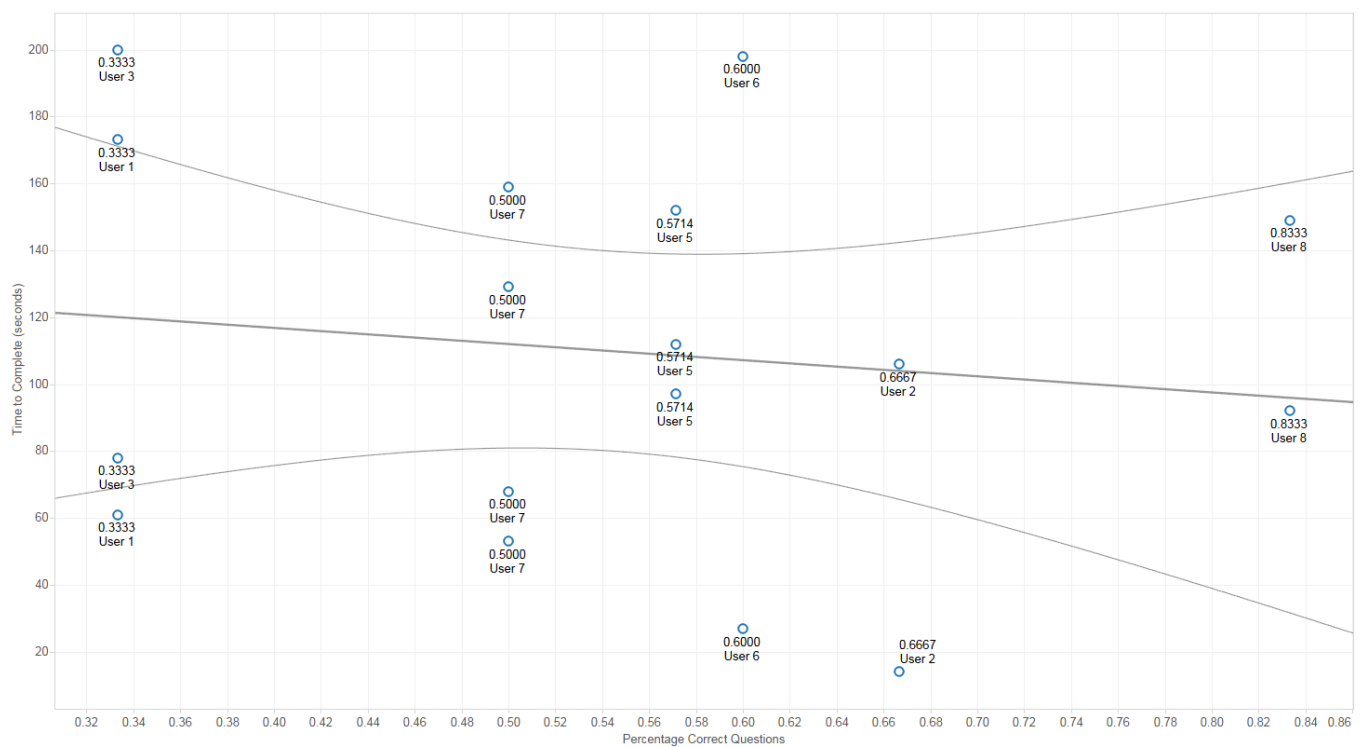
Participant.	Count	Sum	Average	Variance
1	9	368	40.88889	2732.111
2	9	305	33.88889	1242.861
3	10	672	67.2	3915.289
4	12	823	68.58333	2010.447
5	8	735	91.875	7561.554
6	12	601	50.08333	3137.356
7	8	585	73.125	3758.982
8	2	181	90.5	4.5

ANOVA						
Source of Variation	SS	df	MS	F	p-value	F <sub>crit</sub>
Between Groups	23,302.54	7	3328.934	1.017162	0.428226	1.816007
Within Groups	202,911.5	62	3272.766			
Total	226,214	69				

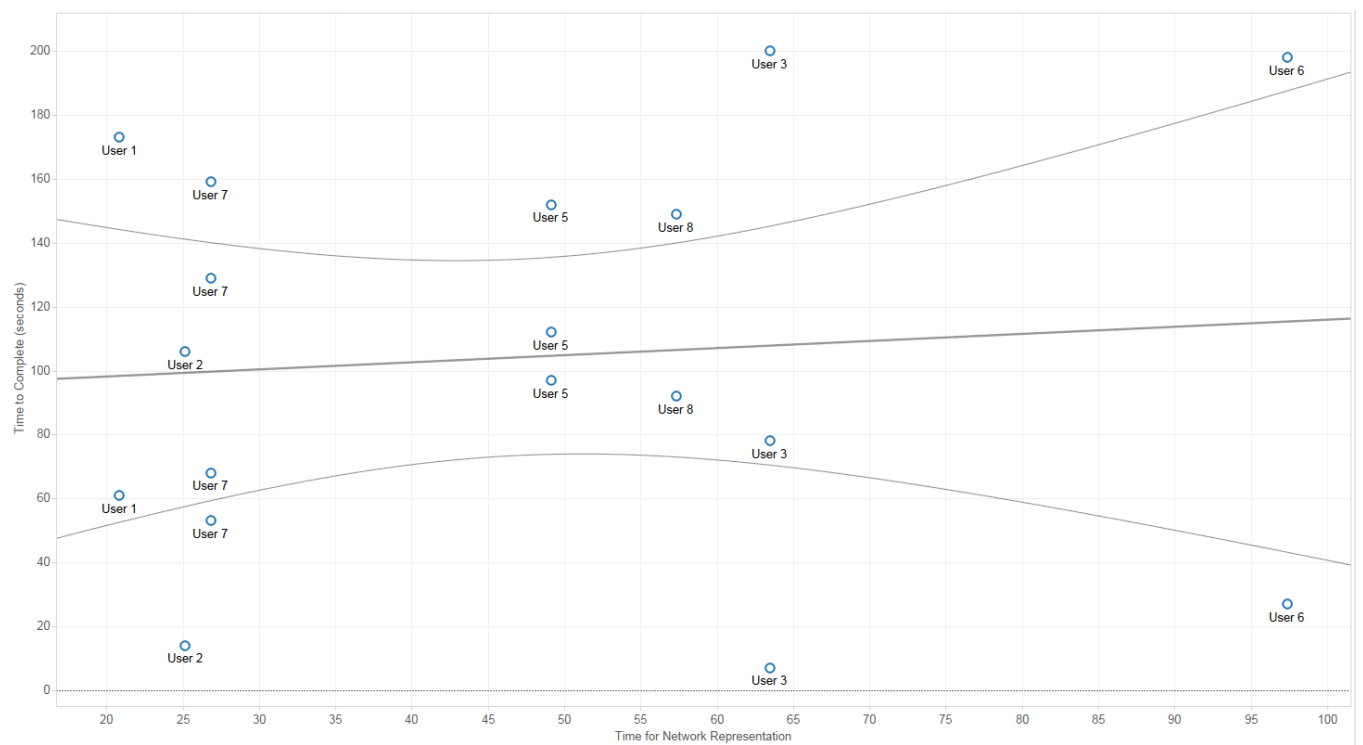
**Table 4.** Results from the ANOVA on users' accuracies with  $\alpha = 0.1$ .

Groups	Count	Sum	Average	Variance		
1	8	4	0.5	0.285714		
2	8	6	0.75	0.214286		
3	8	4	0.5	0.285714		
4	2	2	1	0		
5	10	6	0.6	0.266667		
6	7	5	0.714286	0.238095		
7	11	7	0.636364	0.254545		
8	8	7	0.875	0.125		
ANOVA						
Source of Variation	SS	df	MS	F	p-value	F <sub>crit</sub>
Between Groups	1.138071	7	0.162582	0.688633	0.681104	1.831093
Within Groups	12.74903	54	0.236093			
Total	13.8871	61				

In Figures 11 and 12, we checked whether there was a relationship between either the accuracy (Figure 11) or time spent (Figure 12) on network tasks, and the time spent on other tasks. Intuitively, we see that the figures express no clear relationship. Testing for a linear model resulted in  $R^2 < 0.05$  in either case, suggesting a very poor fit. The same would be obtained for other models such as logarithmic or polynomial. We thus conclude that the time spent in other tasks was not related to what users did with network representation tasks.



**Figure 11.** Relationship between the percentage of network representation questions that a user got correct and the time it took for users to answer the other questions. R-squared for this linear model is 0.017861.



**Figure 12.** Relationship between the time spent on network representation questions and the time it took for users to answer the other questions. R-squared for this linear model is 0.0081678.

#### 4.3. Thematic Analysis

Prior to the usability session, we interviewed eight key informants. A thematic analysis of the interviews reveals five themes, which cover what area of public health the user works in and how they perform their job duties, as well as their current computer solutions and processes. The first theme was the area of work or *portfolio of interest*. Most of our informants work on population-level prevention, rather than individual-level. The main themes were food security, nutrition policy, and tobacco control. These themes were expected as they often appear in the literature on healthy weight policies due to the importance of healthy eating and the inspiration found in tobacco control when it comes to industry regulations. We noted the absence of experts in the well-being sector. Prevalent *use of evidence* was the next theme we found. All but one of our informants mentioned the use of evidence in their jobs. Our informants were consumers of evidence more than creators of it. There was some mention of analysts and surveys of the literature. Informants looked to other health authorities and provinces to see what they are doing and what is working. Not a single informant had a formal process for analyzing the outcomes of their decisions. Our tool may thus be most beneficial to analysts in its current version, or it would have to include access to evidence and linking to others' portfolios to be made more relevant to our informants.

We were interested in the users' *current level of systems thinking* knowledges because the software supports that view of the world. Our findings suggest that users are aware of the many different factors that contribute to mental and physical well-being; however, most do not think at a high systems level in terms of how these aspects relate to each other. This indicates an important discrepancy between what constitutes systems thinking to informants and what parts of systems thinking the software is actually designed to support.

The *Software and Tools Currently Used* thematic area yielded some surprising results. We got a very clear message that the users did not use any special software or tools in their policy works. The software and tools that they did use were "not beyond the typical office stuff", such as Word or Excel.

The last theme that we identified was related to the *potential software uses* in workflow. We noticed that both searching for and sharing information were things that users thought software could help them do better. These tasks were not supported in any special way for the users. Software that could help with the "synthesis of the evidence to show the policy direction we should be taking rather than our current policies" was also mentioned. There was less of a consensus in this theme. In the words of a participant, there was a desire for software that could help "people to systems think and to link across different policy areas ... and gosh I think, if we could do that it would be great". The discussion suggested that users were not so clear as to exactly what they needed but may realize it once they see it.

After the usability session, all but one participant completed the final component of the study. After performing thematic analysis, a total of six themes emerged. The first theme was about *training*. While our tutorials consisted of slides, participants expressed interest in alternative forms of training, ranging from videos to case studies or entire sessions. Training would contribute to equipping users with the skills to fully benefit from the software and it would address two of the most pressing issues: our users may need examples to clarify the reasons for taking a systems approach, and they are very short on time. Indeed, the second theme emphasized that *time* was limited for our users. Decision makers live in a very fast-paced environment and our tool must be usable within these constraints. This raises the challenge of having a tool that is versatile enough to support decision-making processes in a variety of situations, while being specialized enough to be used quickly.

Third, participants discussed what *uses the software could have*. This topic tended to be polarizing, as participants were either thinking of a variety of applications or were not sure about the tool or the paradigm of systems thinking in general. The first group was interested in using the software for exploring, planning, or justifying policies. The software includes a tool to find what a policy will impact, which directly supports the task of ensuring that policies impact that which they are supposed to. It also provides a wealth of information that participants reported using for planning, and it gives a chance



to explore and brainstorm about the system before embarking on a project. On the other hand, the paradigm of systems thinking, and the promises of using it in obesity, have been around for several years. Participants have been well exposed to it through the Foresight Obesity Map, though that first experience may not have been positive. This can lead to considering new initiatives with caution. As aforementioned, case studies could play a role in demonstrating benefits.

The fourth theme was on the *integration of the software* within an existing workflow. Every single participant had a different idea. One needs bullet points on a topic, another would like statistics, a third wants a list of references, and a fourth wishes “to use it in a presentation to show the complexity” (as the Foresight Map which is often used to exemplify the complexity of obesity). We hypothesize that the wide variety in responses results from each participant having a unique workflow. In case the audience truly has a process that varies entirely from person to person, there should be export options in the software to satisfy different needs (bullet points, numbers, references, pictures).

The fifth theme focused on how the software *deals with the evidence*, which was already emphasized in the systems interview as being essential for policymaking. Like any model, the relevance of the map within the software to the real world gradually decreases unless it is maintained. A sustainable way to maintain it would be necessary in order to use it in the long run. In addition, two approaches to the ‘evidence’ emerged. One was concerned about getting the most recent information, whereas another favored the strength of evidence by focusing on systematic reviews. There is thus a choice to be made between having access to the latest advances (e.g., through expert opinions) or the most agreed upon mechanisms (e.g., through systematic reviews).

Finally, participants had an opportunity to *suggest other features*. Few suggestions were provided, perhaps because this came after a long conversation and because it was the first time seeing the software. Suggestions were to use it as an app or seeing a list rather than network.

## 5. Discussion

We observed that users were learning while they were performing the network representation tasks. The questions started out easy and progressed upwards in difficulty. The idea was to be able to see more accurately at what point a user was having trouble. However, what ended up happening was that users would first answer incorrectly and then on subsequent harder questions (which required a closer look at the screen, sometimes literally) they would notice their own mistake, similar to a scaffolded learning experience [67]. Only their initial answer was recorded, but this might not reflect their final understanding of the system or software. This finding may be useful when constructing training material to accompany the software.

The tasks were performed while adhering to the think aloud protocol, which is typical for a usability study, but this may have lengthened the time it took for users to perform a task. The presence of a keen observer is also expected to be atypical.

One of the challenges faced by health and government workers is a lack of time. This prohibits extensive training to be given to participants before testing of the software. If we were to test the software on students, then we could also see how providing training influences the results [52]. Partnering with a university class may also prove fruitful, perhaps a network analysis class, in that the paired analysis approach might also be tested. Paired analysis is a data analysis approach where both a tool expert and a domain subject matter expert sit together and use the same computer. The software expert handles the navigation while the domain expert guides the session by asking questions and identifying what is interesting and relevant to him or her. Such a study may provide further insight into the tool’s usefulness and capabilities without requiring that subject matter experts learn the software during their short study session.

The main limitation in this study stems from the paucity of work dedicated to the design of software to support systems thinking or the evaluation of systems thinking

tools for policymaking. Our study contributes to addressing this gap and will provide a point of reference for researchers who seek to evaluate their tools. However, this gap in the literature implies that we cannot compare our study with closely related works. The findings of our study are the product of three groups of factors (specific software used, specific application domain of obesity, specific set of participants). Hence, we cannot isolate the effect of a particular construct and venture into conjectures about potential effects in a different environment. If further studies design new systems thinking tools and examine their usability in different application domains, it will then be possible to produce more generalizable findings. Although the evidence base does not allow the generalization of our *findings* regarding the relations between systems science and software usability, we believe that the revised *software* can be used in other application domains since it supports a variety of core systems thinking tasks that are not application specific. For example, we have used the same software to recently analyze systems maps of adolescent suicide [2] since the same questions were of interest: what are some of the loops that lock individuals in cycles of harmful behaviors? How would an intervention on societal constructs (e.g., economic support for families) permeate the system and eventually affect suicidal thoughts in individuals?

Further studies could assess whether community members or practitioners may benefit from the software. For example, research suggests that performing a comprehensive assessment offers valuable information when starting an intervention with a patient [68]: if the patient was to use our software, then it could help the practitioner in focusing the assessment. Reducing the gap between a practitioner's conceptualization of obesity and well-being and that of patients has been shown to be highly valuable in designing solutions that guide practitioners in truly addressing their patients' essential needs [69,70].

In the same manner, as the software might benefit the relationship between individuals and their practitioners, it could be used to benefit constituents and policymakers. As policymakers are cautious in what they perceive constituents may endorse [71], this could leverage the crowd [72] to facilitate the identification of areas that policies could more easily tackle.

## 6. Conclusions

Practitioners agree that systems thinking is essential for the design, implementation, and evaluation of complex population health interventions. However, there is a gap between this agreement and the uptake of systems thinking methods. Our usability study shows that seasoned policymakers face several barriers when using a software tool designed for systems thinking. Although they are ultimately able to use the tool for various analyses, this achievement comes at the expense of time. In addition, the uptake of a tool is questionable given the variety of individual workflows. There is thus a greater need to characterize and ultimately standardize (where possible) the software ecosystem used for policymaking while providing additional support (e.g., built-in tutorials and continuing education) for concepts of systems thinking.

**Author Contributions:** Conceptualization, P.J.G.; methodology, P.J.G. and C.X.V.; software, P.J.G.; writing, P.J.G. and C.X.V.; project administration, P.J.G.; funding acquisition, P.J.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** The first author received funding from the Provincial Health Services Authority (PHSA) of British Columbia to conduct the study via a Statement of Work (SoW). The funder was not involved in the scientific conduct of the study (e.g., methods, analysis).

**Informed Consent Statement:** The information and consent form for the user testing sessions was designed and approved by the PHSA. Participants agreed to the consent form and freely consented to participate in the study.

**Data Availability Statement:** The anonymized data presented in this study are available on request from the corresponding author. The data are not publicly available to protect the identity of the participants involved.

**Acknowledgments:** This paper originated as a report to the Population and Public Health Program at the Provincial Health Services Authority (PHSA) of British Columbia. We thank Trish Hunt (formerly Senior Director at the PHSA) and Sarah Gustin (formerly Knowledge Translation & Communications Manager at the PHSA) for their help in identifying participants and authorizing the report to be declassified for publication. Finally, we are indebted to Rama Flarsheim for assisting with the usability study.

**Conflicts of Interest:** The first author was funded to conduct this study by the PHSA. Methods and analyses were chosen in consultation with an independent expert (second author), without influence by the PHSA. The PHSA contributed to editing the report for clarity.

## Appendix A. Key Informants

**Table A1.** List of key informants within British Columbia, Canada. The PHSA and Ministry of Health have a province-wide mandate, whereas other organizations are regional health authorities (VCH, FHA, IHA).

Name	Organization	Position
Megan Oakey, MPH	Vancouver Coastal Health (VCH)	Healthy Communities Coordinator, North Shore
Dr Helena Swinkels	Fraser Health Authority (FHA)	Medical Health Officer (MHO), Lead for Chronic Disease Prevention and Healthier Community Partnerships
Melanie Kurrein, MA, RD	Provincial Health Services Authority (PHSA)	Provincial Manager, Food Security
Shelley Canitz	Ministry of Health	Director, Tobacco Control and Injury Prevention Healthy Living and Health Promotion Branch Population and Public Health Division
Matt Herman, MSc	Ministry of Health	Executive Director Chronic Disease/Injury Prevention and Built Environment
Anna Wren, RD, MPH	Ministry of Health	Project Manager, Office of the Provincial Dietician
Kitty Yung	Vancouver Coastal Health (VCH)	Registered Dietician
Nadine Baerg	Interior Health Authority (IHA)	Public Health Dietician

## Appendix B. Introductory Script to the Session

Hello, [name], thank you again for joining me today . . .

Before we begin, I have some information to share with you, and I'm going to read it to make sure that I cover everything and give every participant the same instructions.

Philippe Giabbanelli along with the PHSA has developed a piece of software that maps how various factors contribute to obesity. Today we would like to evaluate how user-friendly that software solution is. We believe it is vital we get feedback from the potential end-users of this solution. I am going to share the solution with you on my laptop today, and ask you to do some fake tasks on it. After we get feedback from multiple customers, we will make appropriate changes to the solution. So keep in mind that what we launch may not be *exactly* what you see today!

I want to make it clear that we are testing the design of our solution today, not you. You are helping us evaluate the design, and we really appreciate you taking time for this. There are no mistakes and no wrong answers—anything you say and do helps us create the best solution. Don't hold back!

Be yourself and act as you normally would when you use the solution. The only thing I'd like you to do differently is to think out loud. What I mean by that is say what you're looking at, what you're trying to do, what you're thinking. If you don't see something you're looking for, tell me what is missing. For example, you could say something like "I am looking for the Save button. Oh, there it is at the top right corner. I expected it on

the left.” This makes it easier for me to understand what you are doing and what you are thinking.

Lastly, I’d like to record today’s session, so I don’t need to take notes while talking to you. The recording we make today will only be seen internally by people working on this project. Is this ok with you?

Do you have any questions so far?

I’m turning on the recording now.

## Appendix C. Scripts of Questions and Tasks During the Session

### Appendix C.1. Questionnaire Script

Hello,

Have you read the consent for and do you agree to participate today? I will be recording this session if that’s alright with you.

(1) Could you please tell me about your role here?

(2) In your work do you engage directly with policy?

*If yes:* what kind of policy work do you do? Related to what kind of health issues?

*If no:* could you tell me about how your work is influenced by /or influences policy?

(3) When you think about the policy work that you do, do you think about it involving or affecting the larger policy environment?

(a) What does the policy environment look like? How are the different players or pieces relate to each other?

(b) Can you give me a specific example of a policy that you were involved with and how it fits within a larger policy environment?

(4) Could you tell me about a policy that was done in isolation or not considering the larger context?

(5) Could you please describe your work-flow in relation to policy?

(6) What tools do you currently use? Any software?

(7) These tools and software, do they help you think about or understand your policy in relation to the broader context that it is related to and that might be effected by it? [will use previous example given if need be] How so? Can you give a specific example?

(8) Can you think of any tools or software that might help you think about your policy work in relation to this broader context?

(9) If you had these tools or software, how would they be part of your work-flow?

(10) In your realm of expertise, how would you use \$200,000 to improve well-being? What would this project look like?

(11) Is there anything else about policy, the evaluation of potential impacts of a policy, or software that you would like to mention?

### Appendix C.2. Usability Study Script

Ok, now let’s move on to the usability session. What we want to test today is the usability of this software. I want to make it clear that we are NOT testing you today. There are no right or wrong answers here, so please be honest with your thoughts and feelings. Don’t worry, you can’t hurt my feelings or anything like that.

What’s going to happen is I’ll ask you to go through some tasks on the tool and observe where you have trouble using the tool. Act as you normally would if using this tool, but the only thing I’ll ask you to do differently is to think out aloud. What I mean by that is verbalize what is going on in your mind. For example “I’m looking for a help icon. I’m looking at the top-left. I don’t see it there. Oh, there it is at the bottom-right. I did not expect that because . . . ”.

Do you have any questions before we start?

### Network structure questions:

- Find the “Physiology” theme.

- Find factor “emotional eating”.
- How many factors directly influence “emotional eating”?
- What directly influences “emotional eating”?
- How many factors does “emotional eating” influence?
- What does “emotional eating” directly influence?
- Does “depression” influence “emotional eating”?
- Can you find a relationship that you were previously unaware of?

**Multiple pathways:**

- Does “Weight Bias” indirectly influence “Blood pressure”?
- How many pathways are there from “Healthy eating” to “Blood pressure”?
- How does the relationship between “Exercise” and “Blood pressure” compare to the relationship between “Appetite” and “Blood pressure”?

**Unintended consequences:**

- What might be an unintended consequence of changing “Medications”?
- You are considering impacting “Late night TV watching”, what else needs to be considered for it to have a maximum effect on “Healthy eating”?
- What else might you think about while considering a policy on “Public health messaging around thinness”?

**Loops:**

- How many factors are in the path starting at “Insulin resistance”, going through “Cells intake of fatty acids” and a few others factors, ended at “Insulin resistance” again?
- Please describe the relationship between “healthy eating” and “short sleep duration”?

**Exploratory task:**

- Let’s come back to the question of “how would you use \$200,000 to tackle obesity in your realm of expertise? What would this project look like?” Understanding that this tool is not a resource allocation tool, how could you use this map to make that project a success?” (education, presentation, guiding data collection ... )

**Questions about the tutorials:**

- If you were unsure of how to do something with this tool, how would you seek help?

**Additional loop questions, if there is time:**

- Can you find a loop that you were previously unaware of?
- Is this a positive feedback loop or a negative feedback loop?

Ok; thanks you. This concludes the hands-on portion of the software usability session; but now that you’ve used the tool for a bit I’d like to ask you a few final thoughts

**Appendix C.3. “Overall Thoughts” Script**

1. How likely are you to recommend this tool to a colleague (1 = very unlikely, 10 = very likely)
2. How would you explain this tool to a colleague who opened it for the first time?
3. Overall, what are your thoughts on the Map Explorer?
4. What did you like about it and what were your frustrations?
5. Do you think this tool could be useful for you?
6. Who [else] do you think could find this tool useful?



## Appendix D. Complementary Analyses

**Table A2.** Time required to complete each task, and accuracy of answers (when a correct answer existed).

	Identification	Understanding	Evaluation
Network Representation	Find the “Physiology” theme. Time: $15.7 \pm 18$ Accuracy: $0.857 \pm 0.378$		
	Find factor “emotional eating”. Time: $66.3 \pm 32.7$ Accuracy: $0.833 \pm 0.408$		
	How many factors directly influence “emotional eating”? Time: $55.6 \pm 52.9$ Accuracy: $0.571 \pm 0.535$	Does “depression” influence “emotional eating”? Time: $82.3 \pm 70.9$ Accuracy: $0.333 \pm 0.516$	
	What directly influences “emotional eating”? Time: $24 \pm 38.4$ Accuracy: $0.167 \pm 0.408$		
	How many factors does “emotional eating” influence? Time: $25.2 \pm 34.3$ Accuracy: $0.8 \pm 0.447$		
	What does “emotional eating” directly influence? Time: $3.3 \pm 4$ Accuracy: $0 \pm 0$		
Loops	How many factors are in the path starting at “Insulin resistance”, going through “Cells intake of fatty acids” and a few others, ended at “Insulin resistance” again?  Time: N/A STD: N/A Accuracy: N/A STD: N/A	Please describe the relationship between “healthy eating” and “short sleep duration”? Time: 78 STD: N/A Accuracy: 1 STD: N/A	Is this a positive feedback loop or a negative feedback loop? Time: N/A STD: N/A Accuracy: N/A STD: N/A
Multiple Pathways	Does “Weight Bias” indirectly influence “Blood pressure”? Time: $143.4 \pm 46$ Accuracy: 1	How many pathways are there from “Healthy eating” to “Blood pressure”? Time: $56.7 \pm 63$ Accuracy: $1 \pm 0$	How does the relationship between “Exercise” and “Blood pressure” compared to the relationship between “Appetite” and “Blood pressure”? Time: 159 STD: N/A Accuracy: 1 STD: N/A
Secondary and Tertiary Effects	What else might you think about while considering a policy on “Public health messaging around thinness”? Time: $57.3 \pm 45.9$ Accuracy: $0.667 \pm 0.577$		What might be an unintended consequence of changing “Medications”? Time: $90 \pm 38.9$ Accuracy: $1 \pm 0$  You are considering impacting “Late night TV watching”, what else needs to be considered for it to have a maximum effect on “Healthy eating”? Time: 112 STD: N/A Accuracy: 0 STD: N/A

## Appendix E. Usability Recommendations

### Appendix E.1. Improvements in Network Representation (26 Recommendations)

At the identification level, we formulate the following 21 recommendations:

- Having node labels touch their nodes.
- Having the full label name pop up when hovering over the text instead of hovering over the node.
- Reducing full text popup delay time to 0.
- The two level representation of the network (themes vs. factors) was not well understood. Replacing the theme triangles with something that visually suggests that a subnetwork is contained within it would be helpful.
- Double clicking a factor should “zoom in more” and pull up a description of the factor.
- Calling the themes “theme clusters” or “groups” would better indicate that they contain many members.
- The colors of the arrow heads should match the color of the link.
- The color of the links (red, green) shouldn’t be used as colors of themes as well.
- The small +, – signs are too hard to read, consider removing them.
- Users would like to be able to pan the map like in Google maps.
- Clarify in legend if “strong” or a thick line is a strong relationship or strong evidence for relationship.
- The legend has a grey triangle for Thematic area but grey is the color for the food consumption theme only. Use triangles beside theme names instead of squares.
- When you expand or pack a theme the network redraws itself and you lose spatial orientation. Consider redrawing only the part of the graph that is affected.
- Add all line thicknesses to legend.
- Ability to filter or hide the themes or factors to reduce “the amount going on.”
- Add common shortcut keys like ctrl-F to bring up the search feature.
- Add synonyms to the search feature so that searching “binge eating” returns “eating disorders.”
- Search feature should search for any of the words in the phrase not just from the first word.
- The result of searching for an item should result in the target item flashing or continuously changing color until clicked. Or changing the size smoothly from large to small until clicked. Putting a red circle around the item or a big red arrow would also work.
- Include in the search popup window an indication of what change to look for after the search has been completed.
- The search feature could be a blank search box located in the tool and always available instead of generating a popup window.

At the understanding level, we make the following four recommendations:

- Show the number of items within a theme by a number on the triangle or by the number of nodes in a mini-graph glyph inside of the theme triangle.
- The popup tooltip should be consistent between both themes and factors. Displaying the full name is recommended.
- The length of the line does not carry any meaning and it would be clearer if all links were of equal length.
- There could be more than two levels of organization. Ex. Blood pressure could be “under” cardiovascular.

Finally, at the evaluation level, we formulate one recommendation:

- The ability to export the relationships as a list would be helpful for users. They can take the list and do further action on each of the items on the list (e.g., research it further, or formulate policy recommendations).

### Appendix E.2. Improvements in Multiple Pathways (Two Recommendations)

Our two recommendations are at the identification level only. We recommend:

- The “See multiple pathways between two factors” tool should be a single pop up window where you can choose both the start and end nodes at the same time.
- The “See multiple pathways between two factors” tool should also produce the results of swapping the start and end nodes to show both directions at once. This could be shown as two separate graphs.

### Appendix E.3. Improvements in Secondary and Tertiary Effects (Four Recommendations)

- At the identification level, we make the following three recommendations:
- “find the effects of intervening” tool should show that there are further links beyond the level chosen. Maybe by having outgoing links but not showing the nodes or by showing the  $n + 1$  level but at a smaller scale. This would solve the issue of things looking like they “just end”
- For the tree map representation, add some jitter in the Y direction so that the node labels don’t overlap quite so much. Or make labels slanted.
- The “find the effects of intervening” tool should have the centre most factor in the centre of the circles.

We also make one recommendation at the understanding level:

- Having a tool similar to “find the effects of intervening” but instead of starting at a chosen node, we end at the chosen node. This helps answer the question “what can I do to impact obesity, directly and indirectly?”

## Appendix F. Sample Quotes from Participants

### Systems Thinking Interviews (prior to the usability session)

*“a lot of what we do is work with communities, municipalities. Helping create and build policy, social, etc. environments that make the healthy choice the easy choice.”*

*“you know it’s all on behavior change or supportive policies for encouraging healthy behaviors and encouraging everybody to be healthy and active and eat well and not smoke at whatever weight they are”*

*“We could say it off the top of our heads but because we are in health we are an evidence based practice so everything has to be based on evidence”*

### Overall Thoughts (after the usability session)

*“in public health we don’t have a lot of time to fiddle around”*

*“It’s my frustrations with these complex systems things. I’ve seen a few people use them but nothing ever comes of them. so I have a bias against them”*

*“if I could just go to one place it would have all the information that’s sort of my dream”*

*“How does this respond to current evidence? Am I getting the most recent information when I’m opening this up? ”*

## References

- Bryan, C.J. *Rethinking Suicide: Why Prevention Fails, and How We Can do Better*; Oxford University Press: Oxford, UK, 2021.
- Giabbanelli, P.J.; Rice, K.L.; Galgoczy, M.C.; Nataraj, N.; Brown, M.M.; Harper, C.R.; Nguyen, M.D.; Foy, R. Pathways to suicide or collections of vicious cycles? Understanding the complexity of suicide through causal mapping. *Soc. Netw. Anal. Min.* **2022**, *12*, 60. [\[CrossRef\]](#)
- Perry, J. Climate change adaptation in the world's best places: A wicked problem in need of immediate attention. *Landsc. Urban Plan.* **2015**, *133*, 1–11. [\[CrossRef\]](#)
- Finegood, D.T. The importance of systems thinking to address obesity. In *Obesity Treatment and Prevention: New Directions*; Karger Publishers: Basel, Switzerland, 2012; Volume 73, pp. 123–137.
- Parkinson, J.; Dubelaar, C.; Carins, J.; Holden, S.; Newton, F.; Pescud, M. Approaching the wicked problem of obesity: An introduction to the food system compass. *J. Soc. Mark.* **2017**, *7*, 387–404. [\[CrossRef\]](#)
- Heitman, K. Reductionism at the Dawn of Population Health. In *Systems Science and Population Health*; El-Sayed, A.M., Galea, S., Eds.; Oxford University Press: Oxford, UK, 2017; pp. 9–24.
- Chen, H.T. Interfacing theories of program with theories of evaluation for advancing evaluation practice: Reductionism, systems thinking, and pragmatic synthesis. *Eval. Program Plan.* **2016**, *59*, 109–118. [\[CrossRef\]](#)
- Trochim, W.M.; Cabrera, D.A.; Milstein, B.; Gallagher, R.S.; Leischow, S.J. Practical challenges of systems thinking and modeling in public health. *Am. J. Public Health* **2006**, *96*, 538–546. [\[CrossRef\]](#) [\[PubMed\]](#)
- Mooney, S.J. Systems Thinking in Population Health Research and Policy. In *Systems Science and Population Health*; El-Sayed, A.M., Galea, S., Eds.; Oxford University Press: Oxford, UK, 2017; pp. 49–60.
- Grewatsch, S.; Kennedy, S.; Bansal, P. Tackling wicked problems in strategic management with systems thinking. *Strateg. Organ.* **2021**, *2021*, 14761270211038635. [\[CrossRef\]](#)
- de Pinho, H. Generation of Systems Maps: Mapping Complex Systems of Population Health. In *Systems Science and Population Health*; El-Sayed, A.M., Galea, S., Eds.; Oxford University Press: Oxford, UK, 2017; pp. 61–76.
- Barbrook-Johnson, P.; Penn, A.S. Running Systems Mapping Workshops. In *Systems Mapping*; Palgrave Macmillan: Cham, Switzerland, 2022; pp. 145–159.
- Giabbanelli, P.J.; Baniukiewicz, M. Navigating complex systems for policymaking using simple software tools. In *Advanced Data Analytics in Health*; Springer: Cham, Switzerland, 2018; pp. 21–40.
- Giabbanelli, P.J.; Tison, B.; Keith, J. The application of modeling and simulation to public health: Assessing the quality of agent-based models for obesity. *Simul. Model. Pract. Theory* **2021**, *108*, 102268. [\[CrossRef\]](#)
- Naumova, E.N.; Hennessy, E. Presenting models to policymakers: Intention and perception. *J. Public Health Policy* **2018**, *39*, 189–192. [\[CrossRef\]](#)
- Gittelsohn, J.; Novotny, R.; Trude, A.C.B.; Butel, J.; Mikkelsen, B.E. Challenges and lessons learned from multi-level multi-component interventions to prevent and reduce childhood obesity. *Int. J. Environ. Res. Public Health* **2019**, *16*, 30. [\[CrossRef\]](#)
- Nam, C.S.; Ross, A.; Ruggiero, C.; Ferguson, M.; Mui, Y.; Lee, B.Y.; Gittelsohn, J. Process evaluation and lessons learned from engaging local policymakers in the B'More Healthy Communities for Kids trial. *Health Educ. Behav.* **2019**, *46*, 15–23. [\[CrossRef\]](#)
- Seifu, L.; Ruggiero, C.; Ferguson, M.; Mui, Y.; Lee, B.Y.; Gittelsohn, J. Simulation modeling to assist with childhood obesity control: Perceptions of Baltimore City policymakers. *J. Public Health Policy* **2018**, *39*, 173–188. [\[CrossRef\]](#)
- Srivastava, S.; Nambiar, D. Pivoting from systems “thinking” to systems “doing” in health systems—Documenting stakeholder perspectives from Southeast Asia. *Front. Public Health* **2022**, *10*, 2521. [\[CrossRef\]](#)
- Drasic, L.; Giabbanelli, P.J. Exploring the interactions between physical well-being, and obesity. *Can. J. Diabetes* **2015**, *39*, S12–S13. [\[CrossRef\]](#)
- McPherson, K.; Marsh, T.; Brown, M. Foresight report on obesity. *Lancet* **2007**, *370*, 1755. [\[CrossRef\]](#) [\[PubMed\]](#)
- McGlashan, J.; Hayward, J.; Brown, A.; Owen, B.; Millar, L.; Johnstone, M.; Creighton, D.; Allender, S. Comparing complex perspectives on obesity drivers: Action-driven communities and evidence-oriented experts. *Obes. Sci. Pract.* **2018**, *4*, 575–581. [\[CrossRef\]](#) [\[PubMed\]](#)
- Siokou, C.; Morgan, R.; Shiell, A. Group model building: A participatory approach to understanding and acting on systems. *Public Health Res. Pract.* **2014**, *25*, e2511404. [\[CrossRef\]](#)
- Jack, A. Foresight report on obesity—Author's reply. *Lancet* **2007**, *370*, 1755. [\[CrossRef\]](#)
- Allender, S.; Owen, B.; Kuhlberg, J.; Lowe, J.; Nagorcka-Smith, P.; Whelan, J.; Bell, C. A community based systems diagram of obesity causes. *PloS ONE* **2015**, *10*, e0129683. [\[CrossRef\]](#)
- Grant, S.; Soltani Panah, A.; McCosker, A. Weight-Biased Language across 30 Years of Australian News Reporting on Obesity: Associations with Public Health Policy. *Obesities* **2022**, *2*, 103–114. [\[CrossRef\]](#)
- Papoutsis, C.; Shaw, J.; Paparini, S.; Shaw, S. We need to talk about complexity in health research: Findings from a focused ethnography. *Qual. Health Res.* **2021**, *31*, 338–348. [\[CrossRef\]](#)
- Kwamie, A.; Ha, S.; Ghaffar, A. Applied systems thinking: Unlocking theory, evidence and practice for health policy and systems research. *Health Policy Plan.* **2021**, *36*, 1715–1717. [\[CrossRef\]](#)
- Chughtai, S.; Blanchet, K. Systems thinking in public health: A bibliographic contribution to a meta-narrative review. *Health Policy Plan.* **2017**, *32*, 585–594. [\[CrossRef\]](#) [\[PubMed\]](#)

30. Weinerth, K.; Koenig, V.; Brunner, M.; Martin, R. Concept maps: A useful and usable tool for computer-based knowledge assessment? A literature review with a focus on usability. *Comput. Educ.* **2014**, *78*, 201–209. [\[CrossRef\]](#)
31. Giabbanelli, P.J.; Tawfik, A.A.; Wang, B. Designing the next generation of map assessment systems: Open questions and opportunities to automatically assess a student's knowledge as a map. *J. Res. Technol. Educ.* **2023**, *55*, 79–93. [\[CrossRef\]](#)
32. Giabbanelli, P.J.; Tawfik, A.A. Reducing the gap between the conceptual models of students and experts using graph-based adaptive instructional systems. In Proceedings of the International Conference on Human-Computer Interaction, Copenhagen, Denmark, 19–24 July 2020; Springer: Cham, Switzerland, 2020; pp. 538–556.
33. Machado, C.T.; Carvalho, A.A. Concept mapping: Benefits and challenges in higher education. *J. Contin. High. Educ.* **2020**, *68*, 38–53. [\[CrossRef\]](#)
34. Huang, J.; Yu, H.; Marin, E.; Brock, S.; Carden, D.; Davis, T. Physicians' weight loss counseling in two public hospital primary care clinics. *Acad. Med.* **2004**, *79*, 156–161. [\[CrossRef\]](#)
35. Shiffman, S.; Sweeney, C.T.; Pillitteri, J.L.; Sembower, M.A.; Harkins, A.M.; Wadden, T.A. Weight management advice: What do doctors recommend to their patients? *Prev. Med.* **2009**, *49*, 482–486. [\[CrossRef\]](#)
36. Giabbanelli, P.J. Analyzing the complexity of behavioural factors influencing weight in adults. In *Advanced Data Analytics in Health*; Springer: Cham, Switzerland, 2018; pp. 163–181.
37. Verigin, T.; Giabbanelli, P.J.; Davidsen, P.I. Supporting a systems approach to healthy weight interventions in british columbia by modeling weight and well-being. In Proceedings of the 49th Annual Simulation Symposium, Pasadena, CA, USA, 3–6 April 2016; pp. 1–10.
38. Huiberts, I.; Singh, A.; van Lenthe, F.J.; Chinapaw, M.; Collard, D. Evaluation proposal of a national community-based obesity prevention programme: A novel approach considering the complexity perspective. *Int. J. Behav. Nutr. Phys. Act.* **2022**, *19*, 31. [\[CrossRef\]](#)
39. Dubé, L.; Addy, N.A.; Blouin, C.; Drager, N. From policy coherence to 21st century convergence: A whole-of-society paradigm of human and economic development. *Ann. N. Y. Acad. Sci.* **2014**, *1331*, 201–215. [\[CrossRef\]](#)
40. Hammond, R.A.; Dubé, L. A systems science perspective and transdisciplinary models for food and nutrition security. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 12356–12363. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Giabbanelli, P.J.; Crutzen, R. Creating groups with similar expected behavioural response in randomized controlled trials: A fuzzy cognitive map approach. *BMC Med. Res. Methodol.* **2014**, *14*, 130. [\[CrossRef\]](#)
42. Dube, L.; Wolfert, S.; Zimmerman, K.; Yang, N.; Diaz-Lopez, F.; Arvanitis, R.; Sandra Schillo, R.; Hamalova, S.; Nie, J.Y.; Brown, S. Convergence research and innovation digital backbone: Behavioral analytics, artificial intelligence, and digital technologies as bridges between biological, social, and agri-food systems. In *How is Digitalization Affecting Agri-food?* Routledge: Abingdon, UK, 2020; pp. 111–125.
43. Giabbanelli, P.J.; Crutzen, R. Using agent-based models to develop public policy about food behaviours: Future directions and recommendations. *Comput. Math. Methods Med.* **2017**, *2017*, 5742629. [\[CrossRef\]](#)
44. Vermeer, W.H.; Smith, J.D.; Wilensky, U.; Brown, C.H. High-fidelity agent-based modeling to support prevention decision-making: An open science approach. *Prev. Sci.* **2022**, *23*, 832–843. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Carey, G.; Malbon, E.; Carey, N.; Joyce, A.; Crammond, B.; Carey, A. Systems science and systems thinking for public health: A systematic review of the field. *BMJ Open* **2015**, *5*, e009002. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Hong, Q.N.; Bangpan, M.; Stansfield, C.; Kneale, D.; O'Mara-Eves, A.; van Grootel, L.; Thomas, J. Using systems perspectives in evidence synthesis: A methodological mapping review. *Res. Synth. Methods* **2022**, *13*, 667–680. [\[CrossRef\]](#)
47. Luna Pinzon, A.; Stronks, K.; Dijkstra, C.; Renders, C.; Altenburg, T.; den Hertog, K.; Waterlander, W. The ENCOMPASS framework: A practical guide for the evaluation of public health programmes in complex adaptive systems. *Int. J. Behav. Nutr. Phys. Act.* **2022**, *19*, 33. [\[CrossRef\]](#)
48. Malbon, E.; Parkhurst, J. System dynamics modelling and the use of evidence to inform policymaking. *Policy Stud.* **2022**, 1–19. [\[CrossRef\]](#)
49. Hamid, T.K. *Thinking in Circles about Obesity: Applying Systems Thinking to Weight Management*; Springer Science & Business Media: Berlin, Germany, 2009; Volume 202.
50. Axelrod, R. (Ed.) *Structure of Decision: The Cognitive Maps of Political Elites*; Princeton University Press: Princeton, NJ, USA, 2015.
51. Vennix, J.A.; Akkermans, H.A.; Rouwette, E.A. Group model-building to facilitate organizational change: An exploratory study. *Syst. Dyn. Rev. J. Syst. Dyn. Soc.* **1996**, *12*, 39–58. [\[CrossRef\]](#)
52. Giabbanelli, P.J.; Tawfik, A.A. How Perspectives of a System Change Based on Exposure to Positive or Negative Evidence. *Systems* **2021**, *9*, 23. [\[CrossRef\]](#)
53. Meadows, D.H. *Thinking in Systems: A Primer*; Chelsea Green Publishing: Chelsea, VT, USA, 2008.
54. Malhi, L.; Karanfil, Ö.; Merth, T.; Acheson, M.; Palmer, A.; Finegood, D.T. Places to intervene to make complex food systems more healthy, green, fair, and affordable. *J. Hunger. Environ. Nutr.* **2009**, *4*, 466–476. [\[CrossRef\]](#)
55. Johnston, L.M.; Matteson, C.L.; Finegood, D.T. Systems science and obesity policy: A novel framework for analyzing and rethinking population-level planning. *Am. J. Public Health* **2014**, *104*, 1270–1278. [\[CrossRef\]](#)
56. Provincial Health Services Authority. *From Weight to Well-Being: Time for a Shift in Paradigm?* Provincial Health Services Authority: Vancouver, BC, Canada, 2013.



57. Giabbanelli, P.; Flarsheim, R.A.; Vesuvala, C.X.; Drasic, L. Developing technology to support policymakers in taking a systems science approach to obesity and well-being. *Obes. Rev.* **2016**, *17*, 194–195.
58. Riihihaio, S. Usability Testing. *Wiley Handb. Hum. Comput. Interact.* **2018**, *1*, 255–275.
59. Hertzum, M. Usability Testing: A Practitioner’s Guide to Evaluating the User Experience. *Synth. Lect. Hum.-Cent. Inform.* **2020**, *13*, i-105.
60. Gappa, H.; Nordbrock, G.; Mohamad, Y.; Velasco, C.A. Group-based expert walkthroughs to compensate for limited access to target user groups as in case of chronically ill patients. In Proceedings of the International Conference on Computers Helping People with Special Needs, Linz, Austria, 11–13 July 2018; Springer: Cham, Switzerland, 2018; pp. 71–74.
61. Kostakos, V. The big hole in HCI research. *Interactions* **2015**, *22*, 48–51. [[CrossRef](#)]
62. Fan, M.; Lin, J.; Chung, C.; Truong, K.N. Concurrent think-aloud verbalizations and usability problems. *ACM Trans. Comput. -Hum. Interact. (TOCHI)* **2019**, *26*, 1–35. [[CrossRef](#)]
63. Krathwohl, D.R. A revision of Bloom’s taxonomy: An overview. *Theory Into Pract.* **2002**, *41*, 212–218. [[CrossRef](#)]
64. McGrath, M.; Willcutt, W. The creative use of Thinking Maps to embed Blooms’ Taxonomy within teaching, learning and assessment. *EDUCATIO J. Educ.* **2022**, *6*, 346–372. [[CrossRef](#)]
65. Gorman, J. Learning Objectives for Concept Mapping Based on the Complete Bloom’s Taxonomy to Promote Meaningful Learning. In Proceedings of the Eighth International Conference on Concept Mapping, Medellín, Colombia, 26–28 September 2018; Nipmuc Regional High School: Upton, MA, USA, 2018; pp. 1–11.
66. Deep, A.; Pathan, R.; Mitra, R. Comparing Experts’ Systems Thinking Skill Across Contexts. In Proceedings of the 2018 IEEE Tenth International Conference on Technology for Education (T4E), Kuala Lumpur, Malaysia, 8–9 November 2018; IEEE: Piscataway Township, NJ, USA, 2018; pp. 154–157.
67. Hogan, K.; Pressley, M. *Scaffolding Student Learning: Instructional Approaches and Issues*; Brookline Books: Cambridge, MA, USA, 1997.
68. Sharma, A.M.; Padwal, R. Obesity is a sign—over-eating is a symptom: An aetiological framework for the assessment and management of obesity. *Obes. Rev.* **2009**, *11*, 362–370. [[CrossRef](#)]
69. Fraenkel, L.; McGraw, S. What are the essential elements to enable patient participation in medical decision making? *J. Gen. Intern. Med.* **2007**, *22*, 614–619. [[CrossRef](#)]
70. Mead, N.; Bower, P. Patient-centredness: A conceptual framework and review of the empirical literature. *Soc. Sci. Med.* **2000**, *51*, 1087–1110. [[CrossRef](#)] [[PubMed](#)]
71. Giabbanelli, P.J.; Adams, J.; Pillutla, V.S. Feasibility and framing of interventions based on public support: Leveraging text analytics for policymakers. In Proceedings of the Human-Computer Interaction conference, Toronto, ON, Canada, 17–22 July 2016; Springer International Publishing: Cham, Switzerland, 2016.
72. Prpic, J.; Taeihag, A.; Melton, J. Crowdsourcing the policy cycle. In Proceedings of the Collective Intelligence conference, Seoul, Korea, 24–26 September 2014; MIT Center for Collective Intelligence: Madrid, Spain, 2014.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.