Vassar College
Digital Window @ Vassar

Senior Capstone Projects

2019

Understanding intuitive design

Rebecca Andrews
Vassar College

Understanding Intuitive Design

Rebecca Andrews

Vassar College

April 19th, 2019

Senior thesis submitted in partial fulfillment of the requirements for the major in Cognitive Science

First reader: Josh de Leeuw
Second reader: Ken Livingston

Abstract

Though the question of what makes a website "intuitive to use" is frequently explored in the technology industry, some of the particular details of this exploration limits the usefulness of the field's findings. In this thesis, I propose that the addition of scientific theories and methods can help mitigate this problem and improve the methods, and I explore several specific questions related to intuitive design that can help shed light on this overarching question. In particular, I explore whether our mental models remain constant across different task contexts, and I explore the form of the relationship between mental model accuracy and proficiency on a website. I primarily use the card sorting method to help answer these questions, but I also utilize interviewing and some basic usability tests as well. My results suggest that mental models do not change substantially across different task contexts and that mental model accuracy is indeed a strong predictor of performance on the site. Just as importantly, I was also able to illustrate that my proposed method for improving usability does in fact produce websites that are easier to use than the original.

A big question in technological design is understanding what exactly makes a website "intuitive to use." In everyday terms, people typically take this to mean that the site requires limited effort and time to master, and how to use it is in some way obvious. We expect this to be supported by performance metrics indicating that the site can be used successfully without substantial practice, and we expect an underlying explanation for strong performance in the form of some sort of alignment between the site and the person. This notion of alignment is vague, and perhaps that is in part why it is so difficult to construct sites that are intuitive to use in the first place.

The debate surrounding this particular problem is one of the dominant conversations in the technology field because it is probably the most critical part of building a website, even more so than creating something that is simply visually appealing. In this thesis, I hope to explore some of the ways we can answer this question, and in particular, I hope to show how injecting more science into the way that this question is typically explored, user experience research, can help improve our confidence in the scope and meaningfulness of its conclusions. As we proceed, I will also continue to help deepen our understanding of what "intuitive," and particularly the idea of "alignment," really means in the context of web design.

To do this, I will first discuss the current state of user research in order to illustrate why some of its current practices make it difficult to discover the principles underlying intuitive design. Next I will discuss how including scientific theories and methods in user research can help improve these practices, and in addition, how the application of these theories and methods can provide science with useful information in return. Finally, I will use a particular case study to explore some of the specific questions related to intuitive design that are brought up in these earlier sections. By doing so, I will help deepen our understanding of what really makes a website intuitive to use.

An Introduction to User Research

User researchers hope to understand what kinds of design decisions will allow the user to seamlessly transition into using the site. All sorts of design decisions can negatively or positively impact the ease of this transition, and these decisions can range from text content to element placement to color. From the outset, it seems incredibly difficult to determine what values of each will combine to create a website that is intuitive, and yet this is incredibly important if the site is to be successful. Internet users are not the most forgiving; within the first few seconds, most users will form strong opinions about the visual appeal and structure of the site that informs whether to stay there or try somewhere else (Lindgaard, Fernandes, Dudek & Brown, 2006). From the perspective of the user, a site that is difficult to understand might as well not have any use at all.

In the technology industry, this problem has not gone unnoticed, and as a result, more resources have been devoted to understanding how real users use technology than in the early days when products were built without ever talking to the user. In 1983, there were about one thousand people in the world working in something like user research, while today that number is closer to one million (Nielsen, 2017). This is wonderful news for us everyday consumers, but it does not mean that user research is being utilized to its best ability. In particular, user research is missing some of the methodological and theoretical rigor that might make it a more robust version of applied science.

One of the big problems is the use of industry "best practices" rather than conducting new user research or using the results of previous scientific research to inform design decisions. A closer look at most of these best practices reveals that they tend to be more vague than concrete, which makes it hard to rigorously design and evaluate usability in products. For example, let us look at the following tip about the importance of easy navigation on a website, which was taken from a design website:

> Navigation is about how easy it is for people to take action and move around your website. Some tactics for effective navigation include a logical page hierarchy, using bread crumbs, [and] designing clickable buttons (10 Top Principles of Effective Web Design, 2015).

This tip does provide some possible ways to help users navigate your site. However, there is no real information about how to actually execute these suggestions, such as how one might actually create a "logical page hierarchy" or design "clickable buttons", nor is there any information about under what circumstances each tool is the most helpful. This type of information is important to include because otherwise the tips have no real meaningful translation into action. A designer may think that they are creating a logical page hierarchy when in reality the page hierarchy would be completely illogical to a customer. To make matters worse, principles that are this vague are difficult to rigorously test, so we cannot really know if something like designing clickable buttons really improves navigability. Both of these issues create problems for thoroughly improving usability.

Information of this sort spreads around through word of mouth, which means that many of these rules of thumb survive without ever being tested. Most designers consult blogs written by other designers before anything else to get inspiration for their own projects (UserTesting, 2017). By doing so, designers may perpetuate poor usability without realizing it because these older designs may not well tested with real users. In many cases, not a lot of pretesting is done in advance for most parts of the product unless a part is considered especially critical (Chen, 2018). This seems to be the case because many designers are primarily focused on the immediate aesthetic appeal rather than more generally usability, as they think this will be addressed later on when customers are interacting with the product. In fact, this idea aligns well with the most commonly used developmental process for software today: agile development, which is a development process characterized by an emphasis on flexibility and working with customers in real time.

Even when user research most resembles science, it is still missing some of the key ingredients that would help its results be more robust. In particular, this research tends to be conducted without a lot of controls to ensure that it produces meaningful results. Most usability studies involve less than ten people, do not control for confounds, and use primarily qualitative methods of analysis even when quantitative methods might be more useful (Goodman, Kuniavsky, & Moed, 2012; UserTesting, 2017). Many of these experiments are also subject to a whole range of problems surrounding objectivity, such as confirmation bias, because they do not provide controls to prevent it (Capra, 2006; Piernik, 2017). Under these conditions, user research is not treated as real research by all of those who do it. This is a problem because it means that the outcomes of usability studies are not as easily attributable to one factor or another.

A related problem for user research is that the individuals who might be conducting these research projects do not necessarily see research as a prominent part of their job. In the 2017 UX Industry Survey Report, nearly twice as many of the respondents identified their primary job as design as opposed to research (UserTesting, 2017). This seems to reflect a general bias in the industry towards creating aesthetically pleasing products rather than products that are usable the first day they are out in the market. When I spoke to user researchers over this past summer, one of their biggest problems was getting others in the company to take user research seriously and incorporate it into actual business plans (Beuttner, 2018; Hudson, 2018; Vlahovic, 2018). These individuals cited many reasons for this, such as a misunderstanding about what user research is for, a distrust for qualitative methods, and a general belief that there is not time to conduct research in most circumstances. So while companies may be conducting more user research now than ever, this does not mean that they are always really using it to its fullest potential.

This is the current state of user research in industry. While there are many impassioned user researchers, most companies seem to put a higher value on designing the product than they do understanding the customer. Even when efforts are made to understand the customer, it seems to lack some of the critical controls that might make its results more robust, and more often than not, general rules of thumb are used rather than using old or conducting new research. All of these behaviors mean that user research can be improved. Thus despite the increase in attention paid to user research, we have a long way to go before we can really say that we are maximizing its productive potential.

How Science Can Help

One of the most powerful ways that user research can be improved is by taking inspiration from academic study of similar phenomena, like can be seen in fields like Human Computer Interaction, Human Factors, and Cognitive Science. In the most straightforward sense, there are often academic papers that address specific problems in web design directly by testing and making predictions about different theories of design. There are many examples of such research, ranging from comparing navigation methods (Lida, Hull & Pilcher, 2003), fonts (Bernard, Lida, Riley, Hackler, & Janzen, 2002), and menu styles (Chaparro & Bernard, 1999) to evaluating user expectations with eye tracking (Bernard, 2001). There is even research dedicated to comparing the preferences and needs of specific user groups across different categories such as the region, culture, and age (Barber & Badre, 2018; Reinecke & Gajos, 2014; Mead, Spalding, Sit, Meyer & Walker, 1997), and this is just the tip of the iceberg. The conclusions in these papers can be used in real world situations to inform specific design decisions.

In a more general sense though, academic research supplies two particularly useful tools for those in industry to help motivate more usable design. These are conceptual frameworks, which can help guide researchers towards what types of questions to ask to improve usability, and methods, which can be used to both better understand the state of the problem and actually evaluate usability. These two tools will be further explored here.

Conceptual Frameworks

As mentioned earlier, there are many general principles for design but not necessarily clear strategies for utilizing them. This can make it difficult to know where to start when designing or evaluating an existing website; it is hard to identify what questions one should be asking that might help identify problems and then ultimately find design solutions. This is where borrowing conceptual frameworks from academic research can provide a helping hand by supplying an overarching structure to our inquiry. While there are likely many frameworks that might be useful for web design, one that immediately rises to the top is affordance theory, particularly because it is highly relevant to understanding the problem of how to design interfaces that are intuitive, and it has a long (and somewhat complicated) history with technological design.

Affordance theory was introduced to the world in 1979 by James Gibson in his book *Ecological Approach to Perception.* Gibson's intention was to introduce a concept that would help us understand how living things perceive the world, which in his mind meant by directly perceiving opportunities for action. He describes this concept, affordances, in the following way:

> The affordances of the environment are what it *offers* the animal, what it *provides* or *furnishes*, either for good or ill… It implies the complementarity of the animal and the environment (127).

In this quote we can see what makes this particular conceptualization of acting in the world unique, and why it is relevant for building intuitive interfaces. Rather than seeing actions as emerging only from the actor, the affordance distributes the burden for action between the actor and its world. The classic example of an affordance is the "sit-ability" of a chair; in order for a chair to be sit-able, you need the morphology of both the person and the chair to be just right. In this way, neither agent nor chair is purely responsible for what actions are completed in the world. This was strongly in contrast to many of the popular psychological theories of the time, such as behaviorism or cognitivisim, which placed limited focus on the connection between the world and the agent. For behaviorism understanding an agent meant understanding how the agent reacts to rather than interacts with the world. For cognitivism in this time, the specific details of the world were not really emphasized because everything that you needed to understand the agent was in the mind. For affordance theory, action in the world cannot be explained without accounting for the characteristics of both and how they are related. Under this framework, studying human behavior also means studying our world and our relationship to it.

This is a powerful idea, and unsurprisingly, it found its way into domains outside of its original intended domain of perception. For our purposes of understanding intuitive interfaces, it provides a unique way to understand what actions are likely and "natural" in the sense that they are encouraged by the design of the system and the agent.

Nearly ten years after Gibson's book was published, affordance theory was famously introduced into design by Don Norman in his book "The Design of Everyday Things" (1988). Here, the term was used in conjunction with the term "signifiers," which is used to describe characteristics of an environment that encourage certain actions or make them visible. Norman's key point was that it is both the relationship between the agent and the object and the visibility of this relationship that ultimately affects the behavior of the agent

(2013). Unfortunately, there was initially a lot of confusion in the design community about the appropriate use of the term "affordance" as well as criticism that affordances' roots in theories of perception made its extensions contradictory (Bærentsen & Trettvik, 2002; Kaptelinin & Nardi, 2012; McGrenere & Ho, 2000; Norman, 1999). Today in design, affordances tend to be frequently used alongside ideas such as goals and mental states, while originally affordances were meant to exist completely outside internal states (Gibson, 1979). Nonetheless, the fundamental property of affordances, that it is relational rather than only agent driven, is a powerful starting point to help us guide our design decisions for a website. In fact, it gives us a more concrete way to understand what people mean when they expect "alignment" between a user and their tool in the context of intuitive design. Affordances seem to suggest a kind of critical coherence between the agent and the device that supports certain interactions (hopefully the ones we intend). This theory is also accompanied by a whole range of literature that can help us make the idea of coherence more concrete. Thus the theoretical framework provided by affordances can drive us towards more specific questions to help us improve the usability of a site.

We can start first by asking, following in the footsteps of Rex Hartson (2003), why this relationality need only apply to physical and material characteristics. Users of a website have certain conceptualizations of the system, certain goals in mind, and certain constraints about how they can perceive elements on the page. This might bring us to ask how these mental, functional, and perceptional level characteristics are being supported effectively in our site. Specifically, we might ask:

1. What characteristics of the user and system encourage a certain mental model of the interaction?
2. What characteristics of the user and system make some actions useful and some not?
3. What characteristics of the user and system encourages make the system perceivable (in a literal sense) for a user?

Let's dive in to each of these questions. To really understand the first question, one must first have a good grasp on the idea of a "mental model." A mental model is usually understood as a simplified internal model of the world that living things use to help themselves survive and thrive in the world (Johnson-Laird, 1980). These mental models, sometimes abstractly referred to as knowledge structures, can take many forms, whether as extended metaphors, mappings, problem spaces, or relational networks (Goodyear, Tracey, Lichtenberg, & Wampold, 2005; Staggers & Norcio, 1993). What ties all of these knowledge structures together is that they provide information about the characteristics of objects in the world and their relationships that can be used to solve different sorts of problems. For this reason, many user researchers believe that understanding a user's mental model is the key to understanding what motivates their behavior (Norman, 2013). From this perspective, mental models do not just describe the world we live in but help structure the way we understand it and behave in it.

For example, the "time is space" metaphor has two common versions in English that have different implications for problem solving. The two interpretations are the ego-moving metaphor, where you as the agent move through time, and the time-moving metaphor, where time moves towards you. Depending on which version you are using, the statement “the meeting was moved forward two days,” either means that the meeting is now two days later (ego-moving) or two days earlier (time-moving) (Boroditsky, 2000). Thus the choice of metaphor impacts how a current situation is interpreted. Though this is somewhat of a toy example, it does illustrate how the choice of metaphor impacts how we understand the state of the world and make decisions.

The purpose of the second question is to restrict our interest to actions that the users might actually take rather than the set of all physically possible actions. This is one important distinction from the original introduction of affordance theory into design by Norman, who felt that the definition of affordance does imply that all parts of the screen of a digital devices affords clicking, even if there is no button or any other sort of indicator to suggest to the user to do so (Hartson, 2003; Norman, 1999). While technically true, this does not really seem to match the behavior of any real user of such devices. In the real world, people act with purpose and with the intention of completing their goals. Our goals do affect and often direct our behavior, from our academic performance all the way to our ability to lose weight (Ames, 1992; Linde, Jeffery, Finch, Ng, & Rothman, 2004). Because of the critical importance of goals in our decision making strategies, it is much more useful for us to look at the characteristics that affect the utility of certain actions rather than just the possibility of certain actions.

In the context of the ideas like “intuitive” and “alignment”, these first two questions describe a kind of conceptual alignment between the user and the website structure that support ease of use. We hope and expect that our website is structured in such a way that the user’s underlying mental model helps them make correct inferences about the digital world of the site and how to behave in it, especially with respect to their goals. This is particularly important because prediction is considered to be a central player in human behavior by many researchers (Hawkins & Blakeslee, 2008). Thus by structuring our site to match the predictions of the user, we can make their transition into using it more seamless.

Finally, to attempt to answer our last question, we might take a look at what sorts of information academic research in perception has to offer us. In particular, we might ask ourselves what theories about perception have to say about which types of environments are most aligned with humans’ perceptual strengths, and which seem aligned with our weaknesses. After all, an environmental that is hard to perceive is hard to act in, and actions that are difficult are not as likely to be performed or even considered by the user. Requiring difficult actions will decrease the intuitiveness of the interaction because users will expect key actions to have limited barriers. Thus as designers, we want to make sure that the environment we create for our user is as in line with their strengths as possible.

For example, research in visual perception has shown that in a given fixation, people do not absorb all or most of the details in a visual scene (Coltheart, 1999). Instead, people hold onto the critical conceptual information related to the scene, the gist, and some additional details that are only encoded if they receive direct attention (Smith, & Loschky, 2018). This provides two key takeaways for designers of products like websites: It is important to have a well thought out visual gist because this is what will be remembered by the user, and key information must have some way to draw the attention of the user or it may not be absorbed at all. This is just one example of what we could learn from research about perception, but we will leave the discussion here for now.

Now that we have a sense of what is meant by each question, let us think about how we could use them to try to understand a particular website. In fact, these three general questions have some immediate concrete translations into specific questions that we could ask about a specific site. We might ask:

1. Are there structural parallels between the organization of information on my home page and some familiar system that might act as a mental model?
2. Am I making the purpose of a particular page clear so users understand why to read it?
3. Is the font on the button large enough to be readable?

These questions can help guide our initial exploration of the usability of our site by providing some initial direction, and giving us guidance about the best ways to try to answer some of these questions by looking at relevant research. Notice how these questions have a more concrete nature to them than some of the "best practices" described in the earlier section. These questions are directed towards understanding the current site rather than just describing an ideal, and subsequently, they give us a better sense of how to proceed with our site when we identify a problem. For example, rather than just suggesting that designers utilize a "logical page hierarchy" our first question helps a designer identify what features of their system might help make a page hierarchy logical.

These three questions are certainly not the only questions we can produce by digging deeper into affordance theory. We might also take a closer look at what kinds of environments we think are relevant for understanding affordances in the context of web design. In the affordance theory described by Gibson, there are only two key players: the environment and the agent. We might ask ourselves if a digital system, given that it seems to possess both a physical component (something like the screen), and a digital component (something like a website), is best understood as a single environment when thinking about its role in affordances (Hornecker, 2005). Perhaps each individual environment has features that work with the agent to produce behavior. From here, we might ask:

1. What characteristics of the physical environment make certain actions possible/likely?
2. What characteristics of the virtual environment make certain actions possible/likely?
3. Do these environments have sufficient overlap?

Again, let us dig deeper into the proposed guiding questions. Firstly, the first two questions are related but different in important ways. They both ask us to consider what kinds of alignment our website has with our user at a more concrete versus conceptual level by focusing on the details of the construction of these environments. Here, the interest is in how different features of the environment literally support different actions. However, the specific kinds of features we would look at would likely be quite different between the two kinds of environments.

In our discussion of the physical environment, we might include the mouse or track pad, the keyboard, and the screen. But we also might include parts of the wider physical environment, like the surface the computer is placed on and the chair our user is sitting on. What we include matters for what we consider contributing factors for people's behavior, and much of our discussion here can be inspired by discussions in embodied cognition. At its worst, embodied cognition looks a lot like an unnecessary expansion of these features. A very famous example of this is the 2011 experiment showing that leaning to the left leads to smaller estimates of the height of the Eiffel tower, presumably because smaller numbers are on the left of the number line (Eerland, Guadalupe, & Zwaan, 2011). In this example, there is no clear reason why posture would be a meaningful tool to help to solve this problem. But at its best, embodied cognition research demonstrates that there are more resources for solving a problem, and thus interacting with the world, than just what is going on in our head (Wilson, & Golonka, 2013; Wilson, 2002). For us here too, there might be more critical resources than just the monitor that are impacting the behavior of our user, and so the question is what resources might be useful to include.

For example, if we choose a broader definition of environment, we might include the chair and the surface described above when we study how individuals use a digital product. This may mean we discover that people often interact with our product while sitting on the couch with the computer in their lap, suggesting that their angle of view and the types of hand movements that they are easily able to do may be different than the standard user of the computer. These factors may matter for our technological product. On the other hand, we may not find these factors all that useful for understanding behavior if users are normally at their desk. The important thing for us is to be clear on why a particular feature of the physical world actually matters for the users of our product, and when it does, to include this feature in our analysis of user behavior. For the virtual environment, we can ask many of the same questions, but with a different set of items such as the page layout or even the loading time of the website. This would allow us to similarly look for features of the digital world that might be contributing to the behavior of our users, just as we did for the physical world.

The final question highlights the importance of the relationship between the physical and digital environment when we are interacting with the digital object. We want physical actions to be as congruent as possible with the corresponding responses in the digital world. This is important because congruency is more intuitive; interfering inputs from our senses slow down processing, and compatible inputs improving processing. This effect has been demonstrated with a variety of examples, ranging from the famous original Stroop experiments, to interfering the processing of locations using incongruent words, all the way to facilitating perceptual learning with congruent auditory stimuli (Kim, Seitz, & Shams, 2008; MacLeod, 1991; Stroop, 1935). As much as possible, we want our website design to have congruency between the physical and virtual worlds.

Again, these three questions can translate into concrete questions for our particular site, like:

1. Is the screen size too big to allow effective scrolling?
2. Do I make certain actions prominent and/or hidden by their placement on the page?
3. Is it obvious how clicking in this location will affect what the user sees next on the page?

Just like the last set of questions, these questions can help provide a designer with more specific inquiries to understand their site. For example, if a site requires the user to perform a particular action frequently, such as logging into an account, the designer can use the above questions to determine if the action is facilitated by the physical and digital environments. By separating questions about the environment into questions about the digital, the physical, and their connection, we give designers a way to break down the interaction between the user and the environment into smaller chunks that can be more easily prodded and tested. This distinction between physical and digital environments is practically never mentioned at all in best practices, and normally, only the digital environment is mentioned at all. This means that designers might miss out on opportunities to improve usability because they do not understand the types of physical environments that their users are navigating.

As one final example, we want to look specifically at the types of errors people make on our site. Errors are incredibly useful behavioral data about how well our website is aligned with our users, but because different kinds of errors indicate different problems, it is important to understand the underlying source. We might ask ourselves if the current design of our website is accidentally suggesting that certain actions have particular consequences in our system when they do not (Gaver, 1991), or if our website makes it difficult to execute actions appropriately (Norman, 2013). This might lead us to ask:

1. Is the error a slip (the user seems to understand the system but nonetheless performs the wrong action) or a mistake (the user misunderstands the system so subsequently acts incorrectly)?
2. What characteristics of the user and the system seem to be encouraging either of these types of errors?

These questions give us a framework for understanding errors that occur in our system, and probably more importantly, help us see what next steps we might need to take to decrease the likelihood of the error. When an error is a slip, it may mean something about our current design makes it hard for people to execute their intended action. Errors like these may not require any substantial structural changes because they do not indicate complete confusion about our system, just a difficulty acting in it. These errors are frequently caused by lapses in memory or attention (Reason, 1995). This may require us to look more carefully at things like layout or placement of items rather than organizing principles to help draw attention to the right places and lower the memory burden, or we can also make our tasks require less of these things. By contrast, a mistake indicates that the user is not on the same page about the key principles governing our site. This may mean we might require more substantial changes to help get our website and user to become more aligned.

Just like in the previous examples, these questions can translate into specific questions for a given site:

1. When the user clicked on log out button only to immediately re-login, did they do so because a) they accidentally clicked the wrong button or b) because they thought that the log out button had a different function?
2. In the case of (a), are the buttons large enough to be easily clicked? In the case of (b), is the icon sufficiently clear to indicate to the user that it will log them out?

Questions like these are useful because they will help direct the designer to understand the kind of error made by the user and the larger source of this error. Again, these questions provide the designer with a much more concrete plan of attack than general best practices. In regards to errors, the best practices used in technological design do not specify how to fix them or why the current design is a problem, but simply highlight what the current website is lacking. Revisiting our navigation example, a best practice might propose designing clickable buttons in response to difficulty with a button related action, while our two questions force the designer to think more critically about the source of the error. This might allow them to understand more deeply why a particular button is not successfully clicked.

The point of these examples is to illustrate the real utility of research concepts in evaluating and improving usability. By starting with the relational aspect of affordances as our basis, we were able to pull in additional ideas from related research to create a list of concrete questions we might use to understand the design our website, and all of these questions can help make our site more transparent to the user. Moreover, these specific questions allowed us to dig deeper into the kinds of alignment there can be between a user and a website. This alignment can be purely at the conceptual level (like we saw with the questions related to mental models), at a mostly physical level (like we saw with questions related to physical environment), or somewhere in between. All of these different kinds of alignment support the intuitiveness of a particular website design in different ways, and understanding their differences can help us getter a better sense of how to improve a particular website and websites in general.

The next step after formulating these questions, then, would be to determine how we can best answer them using scientific methods of inquiry. Academic research can also help us in this regard.

Scientific Methods of Inquiry

Academic research also provides industry with a set of well-tested and well thought out set of methods for trying to answer particular usability questions. In particular, we will look at one such method known as "concept mapping" because it is closely related to the overarching research question. Concept mapping is a research technique that involves interacting with a participant such that in the end, you are able to produce some sort of analyzable representation of the mental model, or knowledge structures, underlying a particular set of concepts (Goodman, et al. 2012). As I mentioned earlier, mental models take many forms, and depending on the specific set of concepts or situation at hand, these different knowledge structures are thought to underlie our conceptualization of situations and relationships. Subsequently, there are a variety of different methodological instantiations of concept mapping that range from interviewing, to calculating word co-occurrence, to card sorting (Gonzalez & Juarez, 2013; Goodman et al. 2012; Goodyear et al. 2005). In industry a variety of these methods are actually used, but one of the most common is card sorting.

In a card sorting task, participants are asked to sort a set of concepts, often written on cards, into groups based on their own personal feelings about the relationship between concepts. The basic idea is that if we get a better grasp on the underlying knowledge structure of our users, we can design intuitive interfaces by matching the structure of website or application to that of the person's mental model. In particular, card sorting in this context focuses on a very conceptual level alignment between user and tool. Because this method's purpose is directly related to building intuitive designs, it is a perfect candidate for us to explore further.

Industry card sorting was actually repurposed from studies of psychology, where researchers were similarly interested in understanding how people mentally organize knowledge (Wood & Wood, 2018). The purposes are similar, but in industry, the end goal is to apply this new understanding to construct better interfaces rather than just the understanding in itself.

Card sorting has several key benefits related to its use in technological design. In particular, the result of each card sort provides us with a set of groups produced by each participant. These groups are not unlike a set of submenus that you might see in a website navigation bar which allows us to translate our results meaningfully into structural changes on the site. In this way, the very purpose of card sorting is to support the building of a strong conceptual alignment between the user's mental model and the website's structure, which will hopefully make using the site intuitive for the user. Moreover, there are methods that exist for measuring the similarity between a user's card sort and a website structure, which allows us to quantify the alignment between user and website (Deibel, Anderson, & Anderson, 2005).

In addition, when we use the right analysis strategy, we can interpret the results of a card sorting experiment with confidence that conclusions are unbiased. A challenge for many of the qualitative concept mapping methods is that it can be difficult to account for research subjectivity, which can maker conclusions easier to criticize and results harder to replicate (Conceição, Samuel, & Biniecki, 2017). By contrast, there exist several methods for analyzing card sorting data that allow you to produce aggregate groups directly from the raw participant data before allowing researchers to add in their own qualitative analyses. Card sorting can also be conducted online or in person, with little meaningful difference between the mediums, which makes it much easier for researchers to get a large participant pool (Petrie, Power, Cairns, & Seneler, 2011). This is critical for creating the type of experiments, both in the scientific lab and out in the world, that have meaningfully generalizable conclusions.

However, the industry application of card sorting often misses some of the key principles of the scientific method that help give us confidence in the results of scientific studies. In particular, industry card sorting practices do not really discuss issues of reproducibility and avoiding confounds, both of which have important implications for its results. For example, though there are analytical methods available to assess card sorting data, many researchers opt for a more eye-balling method of analysis, where researchers come together and discuss the larger patterns they witnessed in the sorts and build an aggregate set of groups from there (Goodman, Kuniavsky, & Moed, 2012; Wood & Wood, 2018). Though this type of qualitative analysis has some value, it is not fully substitutable for more technical analyses, particularly because it can be difficult for the human eye to capture some of the complicated patterns in the groupings across many participants. Moreover, subjective discussion early on in the data analysis process leaves room for researchers to inject their own personal biases into their results in ways that hide the real patterns in the data. This could easily contaminate the end groups they create.

In addition, very few details of the card sorting task are consistently pretested to ensure that they will not confound the final groupings. Often, one to few word phrases are used on the cards rather than full descriptions, which leaves room for misunderstandings or divergent interpretations on the part of participants (Wood & Wood, 2008). This is another place where implicit bias about concepts structure can creep in, as the choice of what word(s) to place on the card can make some similarities obvious while obscuring others. While in the research setting such stimuli might be subjected to an entire experiment solely to ensure their objectivity, in industry, they might be compiled by only a few individuals with little emphasis on the objectivity of the description.

These sorts of challenges can be addressed by taking inspiration from efforts in science to improve its own experimental methods. In recent years, greater emphasis has been put on understanding and controlling for "researcher degrees of freedom" or design choices made throughout the experimental process that can affect the outcome (Wicherts et al, 2016). The increased focus on controlling research degrees of freedom has helped illuminate what sorts of studied phenomena have robust evidence and which do not (Andrews, de Leeuw, Larson, & Xu, 2017; Simmons & Simonsohn, 2017). Thus while not all science is done perfectly, greater attention to these details has made a difference and applying the same standard to industry card sorting is likely to have similar benefits.

In order to decide the best way to use and improve a particular method, it is important to recognize what baseline assumptions underlie it. This is particularly important here because the goal of this method is to make intuitive website structures, and we wish to understand what factors contribute to this. For card sorting, two assumptions in particular stand out. The first is that no matter the task you are performing on the site, your mental model of the relationships between content remains constant. This is important because card sorting tasks are presented to the participant as an opportunity to structure information related to the website in a very general way, so it does not account for differences between tasks. This might mean that the concept map formed by participants is not the one they use when faced with particular tasks, and importantly, maybe not with the tasks that are most critical to the site. Connecting this back to our overarching question about intuitiveness, it might be that what is intuitive in one context may not be in another similar task context.

In fact, there is a wide range of research suggesting that for the most part, people's conceptualizations of categories and relationships do vary by situation. Depending on the direct context that people are provided, they will change how similar they rate disparate objects, and many researchers even believe that categories are dynamically constructed each time they are needed (Barsalou, 1982; Barsalou, 2002; Casasanto, & Lupyan, 2015). Specific to product research, there is even evidence that both short term and long term goals can affect similarity ratings as well as even somewhat minor changes in context, suggesting that our intentions might substantially affect how we create groups (Ratneshwar, Barsalou, Pechmann, & Moore, 2001; Ratneshwar, & Shocker, 1991). This seems to bring the earlier assumption under question, as it is possible that researchers and industry workers alike are assuming incorrectly that asking people to describe their concept maps in general will be useful in any specific case. For industry, this could mean wasted dollars and unusable products because the mental model being used during critical tasks on the site may be different than the site structure that is created during a typical card sorting task. Fundamentally, they would be misunderstanding what it means for any given website structure to be intuitive for a user: The attribute of being "intuitive" might not a context independent trait for a given website but rather depends on the task at hand.

For research, assuming that mental models are the predominantly static may mean the many other theories built on top of this idea to also draw incorrect conclusions.

The second critical assumption that underlies card sorting is that the closer you match the website structure to a person's mental model, the easier it is for them to navigate the site. This is the exact reason that it is useful to use card sorting to improve your website in the first place; you want your website structure to match your customers' as closely as possible. If this assumption is true, navigating the website is intuitive precisely when it is highly similar to the users' own mental model. The potential problem with this idea is that while people may have relatively complicated networks of concepts in the abstract, in practice, people may not have the time nor computational power to fully utilize this same network to solve problems. In real world tasks, people are often content to work fast and approximately even if they have the ability to work slower and more deliberately. This observation has sparked ideas about human beings having distinct problem solving strategies, often called "System 1," a set of strategies meant to be automatic and fast, and "System 2," a set of strategies meant to be as accurate as possible at the sacrifice of speed and energy (Kahneman, 2015). Additional research has shown that under many different circumstances, ranging from decisions about probabilities to making basic choices between two things, people often use approximate decision making methods (heuristics) as their primary tool instead of more complicated and accurate strategies (Gigerenzer, Todd, & ABC Research Group, 2001; Tversky, & Kahneman, 1974). This seems to suggest that fully fleshed out mental models may not be called upon in these situations.

A great example of this is research on change blindness, a phenomenon characterized by the lack of awareness of changes in an environment, even when those changes are blatant (Rensink, 2005). This lack of awareness is usually understood to be the result of incomplete modeling of the world; rather than capturing all the specific details of the scene, something more approximate is used instead (Simons, 2000). Examples like this illustrate that humans may not always rely on fully fleshed out mental models when confronted with real world situations. Perhaps, like for visual scene perception, when we navigate a website structure use something simpler and approximate rather than a complete mental model of the relationships between content related to the site's topic. In the industry setting, this may mean that more complicated website structures are used than is optimal for best performance. In the research setting, this may provide greater evidence for theories of problem solving involving bounded rationality. Exploring whether this assumption is true will thus mean something for both industry and science.

From discussing some of the theoretical assumptions that card sorting relies on, we can see that even some of the most critical assumptions of the method are not completely untouchable. Exploring these assumptions further will have benefits that are two-fold: their accuracy has important repercussions for both how the card sorting method should be applied and the resulting website structures that the method creates, and for other scientific research that uses these assumptions as a base. Here lies a perfect opportunity to use a rigorous and scientific application of the method to give back to the theory. If we formulate the right types of experiments that involve manipulating card sorting, we will be able to see if the results indicate these original assumptions are valid. Moreover, these types of questions are directly relevant for understanding what kinds of characteristics are contributing to the "intuitiveness" of a particular website design.

Case Study: Vassar Career Development Office

So far I have discussed the problems related to user research in industry, the contribution science can have to help address these problems, and several theoretical questions related to the idea of "intuitiveness" that can be explored by looking closely at their applications. Now, I will pull this all together with a singular case study website. In the following sections, I will introduce you to the case study website, I will utilize scientific frameworks to help determine potential problem areas, and I will explore ways to improve this website that will also help us understand what factors affect this improvement. In this way, the case study will be a sort of proof of concept for the ideas described in the introduction.

Let me first get you acquainted with the case study website, the Vassar College career development office (CDO) website, located at https://careers.vassar.edu/. The two primary purposes of this site are to support younger Vassar students in finding potential career paths and summer working opportunities, and to help older students find employment and additional higher education opportunities after their departure from the college. An image of the home screen for this site can be found below (Figure 1).

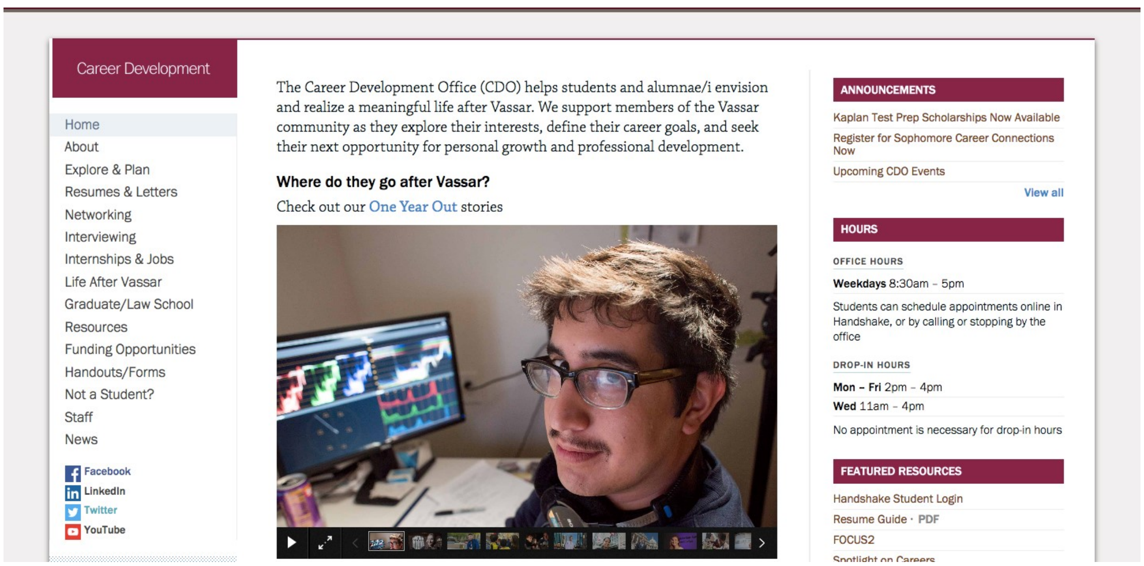

Figure 1. A screenshot of the front page of the career development office website in November 2018. The site is reachable at the following link: https://careers.vassar.edu/

The site is organized into different sections that are indicated by the menu items on the far left side of the page. Several of these menu items lead only to a single web page with no subpages, but most of the menu items have at least 1 sublevel, some with another sublevel beneath that. There are also two additional organizational menus on the right side of the page, which highlight upcoming events and/or deadlines and key resources for students.

The submenus in the left navigation bar mostly correspond to different buckets of actions students might want to complete. For example, the "Explore and Plan" menu item contains several subpages that are meant to help students find majors and careers that are good fits for their interests. The idea behind this organizational structure is that students come to the website with particular goals in mind, and these menu items should line up intuitively with these goals.

One of the interesting problems that this creates in the current design is that some of these goals are highly overlapping, so it is not always clear which submenu will lead to the desired webpage and which ones will not. For example, a student might come to the site wanting to learn what careers people have after graduating from Vassar with a particular major, and resources to help with this problem can be found in the "Explore and Plan," "Networking," "Internships & Jobs," and "Life After Vassar" sections of the site. This means that there are many different places to find all of this related information. On the other side of things, information about job interviews can be found in the "Interviewing" section of the site, but not in the "Internships and Jobs" section, which might be another reasonable place to look. This means that a student may not find information that they expect to be in a particular section when it is in fact available on the site. The fact that information can both reappear in some sections and be divided among others makes it especially important for users to understand what goals each of the menu items is intending to address.

In addition to the majority of menu items that are aligned with student goals, there are a few extra menu items that serve other purposes. Of particular interest are the "Handouts/Form" section, which just contains all the PDFs available on the site, and the "Not a Student?" section, which contains information for alumni, faculty parents, and employers. These menu items are a clear deviation from the overall organizational pattern in that they, in the first case, act as a catchall section, or in the second case, are for the purpose of a different user. For this reason, I will not focus on these sections in the rest of our analysis.

Evaluation of Current Usability of the Website

In order to initially evaluate the usability of the site and to understand the primary users and website curators, interviews were conducted with both staff members from the career development office and Vassar students. In total, eight interviews were conducted with the staff and 20 interviews with students (5 per grade) using a convenience sample. In general, the purpose of the questions directed at the CDO staff members was to get an "insider perspective" on the layout of the site and understand how staff members viewed student needs and usage patterns. The questions directed at students were to give us a sense of how students actually viewed and used the site. The full list of the example questions and a summary of all of the answers can be found in the appendix (Appendix B and Appendix C respectively). In addition, these results were compared to the Google analytics data available for the career development office site from January 1st to December 1st of 2018.

One of the most promising results of our usability evaluation was that staff members had a fairly good sense of what resources students frequently use and consider the most helpful. Eight of the ten most popularly visited pages according to the Google Analytics data, excluding the homepage, appeared in the list of common student uses created by staff. This also match the self-reports from students during the interviews, as five of the six different reasons that students in our survey used the website appeared in this same list complied by staff. In fact, the top two (the Resume Guide and internship/job searching) were exactly the same for both groups. Of the 6 resources considered most helpful by students overall (see Table 1), 5 appeared in some form or another in the list of most important resources for students made by staff (see Appendix C for the complete list). This suggests that staff members do have a good sense for what information should be emphasized to help students out the most, and as a direct result of this, many of these resources are present in the "Featured Resources" sidebar that appears on every page in the site.

Table 1.

*Top Resources Students Cited as Most Helpful**

1. Internship and job listings 2. Resume help 3. Service to match internships/jobs to students based on interest 4. Specific examples of what students with their major do after Vassar 5. Resources to connect to alumni who have career paths of interest 6. Interview help
*The top six helpful resources are included here rather than the whole set because upwards of 30 different resources were suggested by students, with many resources being highly similar or overlapping. All the resources listed in this group were mentioned by at least 5 students in the course of the interviews, and are in order of most frequently mentioned.

However, my interviews did reveal that there were some important disconnects between what problems students thought the CDO website was designed to address and the problems it was actually capable of addressing. In general, students saw the CDO as being primarily focused on the process of finding and securing a job or internships. This meant that they did not recognize the role it could play in steps earlier or later in the process, such as helping the students find careers of interest or determining how to make an internship/job experience economically possible once already secured.

In particular, underclassmen tended to see the website (and the CDO itself) as being useful only once the student already had a sense of what type of career path they were interested in pursuing. In fact, of all the underclassmen interviewed, none of them mentioned wanting help finding potential career paths based on their major on the site in the first pass. Only when students were prodded, (in particular, when they were asked, "Do you think you have a good sense of all the things you could do with your major?") did students recognize that this might be a resource they would find helpful. Often, this sparked a longer conversation about the different kinds of resources this could be (speaking with alumnae/i, quizzes, etc.). For these students, it was not immediately obvious that the CDO could help them with this problem.

The Google analytics data seems to confirm this same pattern. Through all of 2018, the explore and plan section header page, which is the opening page for a section dedicated to helping underclassmen find careers or majors of interest, was only the 11th most popular page on the site, and the 8th most popular subsection header (out of 15). This heavily contrasts with the student created list (Table 1), which puts career exploration as the third and fourth most helpful type of resource. This suggests that while students recognize the importance of this problem for themselves, it is not immediately clear to them that the CDO would be the venue for them to address it.

This problem was not just isolated in underclassmen, but was prevalent across grades. One such example of this would be that funding for summer internships was referenced as a somewhat helpful resource (four students mentioned in particular), and was cited as the most helpful resource of all for students by staff members, and yet not a single person interviewed used the website to secure funding. The Google analytics data seems to tell a slightly more complicated story, as the internship funding page was the second most popular page on the site, but the funding subsection header page, which would normally be navigated to before the internship funding page, was only viewed one third as many times. This seems to suggest that most of the students who get on the internship funding page are being directed straight there from some other source rather than simply navigating the site. None of the other funding subpages make even the top ten in the list of popular pages, with two pages of the four funding pages not even making the top twenty. Though students are at least able to get to a key funding resource, they do not seem to be fully utilizing the other resources in this section.

I also noticed that students tended to suggest different organizational principles than those used on the site. During the interviews, students were asked what features they would like in a career development website designed for them specifically. More than half the students (fourteen of the twenty) mentioned specific ideas related to the site structure, but only one student described a site structure based first and foremost on task, which is the organizational principle used now by the site. The most commonly described organizational pattern was major/interest/discipline, though students varied on whether or not they included all three in their description. In total, seven students mentioned this organizational pattern. The current site has some underlying structure related to this (such as separation between graduate and law school), but it is not the baseline organizational principle. This seems to suggest that students do not necessarily structure this information in the way that the current website does.

A related problem I discovered through the course of my interviews was that many students had very limited or no experience with the CDO website, even of those who had made appointments and utilized the office. One senior, when asked why they use Google instead of the career development office website for things like resume tips, responded that they "did not even know we had a website." Of particular concern, nearly all of the underclassmen that were interviewed were completely unaware that there was a website even if they were aware of the office. Though this was not a completely random sample, there is no clear reason to suspect that the convenience method utilized would be more likely to select students with less than average experience with the CDO website. This seems to suggest that there might be many students who are not even aware that this exists as a resource for them to use.

In general, the fundamental problem here seems to be a misconception about what information is actually available on the site. This seems to indicate a kind of conceptual misalignment between the student users and the site's structure, which is supported by different ideas about how the site should be structured and misunderstandings about what kinds of resources are available on it. While a particularly large part of the problem is that some students never make it on to the site at all, it is quite alarming that even of those who do, many do not have a clear picture of what resources they have at their disposal. Thus while it is outside the scope of this thesis to determine what strategies can be used to get students to try the website, we may be able to apply some of the strategies described in earlier sections to make it more obvious to students what resources are available when they first land on the site. It is clear that this site, like many, is still trying to understand how to make its usage intuitive for its key user group.

Applying What We've Learned to the Case Study

I have described in the previous section the state of the career development office website and proposed a potential problem, a misunderstanding of what information is available on the site, that needs to be addressed. Now, I can take the theoretical framework described earlier, affordance theory, and use it to help us understand how this problem might be addressed. As a reminder, affordance theory asks us to think about how features of both the tool and the user relate to one another to produce (or to not produce) action. In the previous section, I produced many example questions to help guide our exploration of a site's usability and the kinds of alignment it may or may not have with its users. Given the particular nature of the problem for the site, which is that it seems to be at least partially related to conceptual misalignment, it makes sense to begin with these kinds of questions. Let's pull on the first question to see if this can help get us started.

1. What characteristics of the user and system encourage a certain mental model of the interaction?

This question seems particularly relevant because the problem seems mostly related to a misunderstanding about the network of information on the site. Looking at the case website, there does not seem to be an obvious parallel between it and a simple other physically system, like there might be between a grocery website and a grocery store (such as organizing submenus by food category), that a user could use to help build a mental model. It seems unlikely that people have enough experience navigating a career development office space or that there are even any consistent organizational principles of these offices that could be extrapolated off anyway as well. However, that does not mean there is not some underlying reason, like an organizing principle, that students might use to organize career development related information. In the interviews, many students proposed organizing the website around majors or industries, and others did not have any clear principles at all. Perhaps then, it makes sense to explore how students do conceptualize the relationships between the types of information available on the site. If they seem to use different organizational principles, this may explain why students are not finding and using certain helpful resources.

Fortunately, I have already discussed a particular method that can be used to help understand how people organize information: card sorting. By creating a set of cards containing descriptions of the content pages of the career development office site, we could see how students would organize this content themselves. This would give us a better understanding of the mental tools they are using to navigate the site, and moreover, we would ultimately be able to use their groupings to inform a restructuring of the website content. This would allow us to improve the ease by which students navigate the site by improving the mapping between our site and their mental model.

But we do not have to stop here. As I mentioned earlier, it is possible that there are several assumptions underlying the normal application of card sorting that will turn out to be incorrect. These assumptions are related to the kinds of features that actually systematically affect the conceptual alignment between users and a website.

It may be the case that people form different groups under different goal-oriented contexts. We could check this assumption by trying several versions of the card-sorting task, each involving either highlighting a different task that might be completed on the site, or no task at all. If the same groups are formed under all the conditions, we could infer that the mental model underlying our usage of the site is not changing between contexts, indicating that there is a salient base mental model. If we saw different groups formed under the various conditions, we could infer that mental models are indeed impacted in important ways by goals, even between similar types of tasks. This would give us a sense of whether when we seek to improve the "intuitiveness" of a current website, we need to be thinking about particular task contexts. This would force us to consider what tasks are most important on our site and highlight those when we ask participants to make groups or if not. These kinds of decisions have important implications for how we might conduct card-sorting tasks in the future to maximize the ease in navigating our site.

It is also possible that there is not a clear positive relationship between mapping of mental model to site structure and performance on the site. Evaluating this relationship could give us a sense of whether mental model similarity is directly related how intuitive the user finds the website. We could check this assumption by asking participants to create groupings in the card-sorting task and asking them to navigate the site, and then determine if there is a strong relationship between matching mental model and site structure and performance. If there is, this would suggest that subjects may be directly using their mental model to help direct their behavior. If there is not, this might indicate a more complicated relationship between mental model and its actual usage in goal oriented situations. Revealing which of these conclusions has more evidence would allow us to understand how the knowledge of card sorting can best be applied. As of now, we often try to take the final groups from the card sorting task and make our website structure as similar to this as possible. This experiment may reveal that this is not the optimal way to structure the site.

In summary, we could explore the two following theoretical questions, which are importantly related to intuitive design while attempting to improve the usability of the case website:

1. To what extent are mental models invariant across tasks?

2. Does increasing the similarity of a system to users' mental models improve their performance in a similar way?

If we go about answering these two questions, we would not only shed some light onto important theoretical assumptions believed by many scientists, but also in the process, reveal ways that we could improve the application of the method that may help us produce more intuitive interfaces. In this way, we would see a symbiotic relationship between academia and industry. Working to understand and improve one can subsequently do the same for the other.

There are other questions brought up earlier in the introduction that we might ask about the career development office website, but the two questions above seem like enough to go on for now. We have an idea of how these questions relate to the larger usability problem, we have an idea of how we might try to answer them, and we have an idea of some of the larger implications of finding the answers. The next step is simply to do it.

Two Experiments to Answer Two Key Questions

In order to evaluate the two theoretical questions proposed in the introduction, I conducted two separate experiments involving the card sorting task. The main purpose of the first experiment was to understand if the users' mental models would remain the same under different task contexts. This was the first theoretical question I set out to answer. The larger question underlying this is whether our mental models shift when we use them in different task contexts, even when these contexts are relatively similar. The main purpose of the second experiment was to explore the exact relationship between mental model and system similarity and fluency using the system. This was the second theoretical question I set out to answer. The larger question here is we truly utilize our mental model directly in its full form when we navigate the world or if the relationship between mental model and task performance is more complicated. All of the resources used to conduct these experiments can be found at https://osf.io/gd64c/.

Part 1: Importance of Task Context

In this experiment, I sought to explore to what extent mental models are invariant across tasks. In particular, I hoped to see if highlighting the different contexts one might use the Vassar career development website would change how students organized their groups. Based on earlier research involving dynamic mental models (Ratneshwar, & Shocker, 1991; Ratneshwar, et al., 2001), we should observe different mental models, i.e. different created groups, in the different task contexts. In particular, we might expect that concepts that are the most highlighted in a given task context will be broken down into further subgroups than concepts that are not clearly mentioned in the context.

Method

Participants

127 participants were recruited from Vassar College and 75 participants were recruited from Prolific, an online resource for participant recruitment. All participation was voluntary, and all Prolific participants were compensated for their time. Participants were all self-identified undergraduate students, and those from Prolific could be from any undergraduate institution across the globe. In total, there were 55 first years, 50 second years, 37 third years, 47 fourth years, 5 fifth years or more, and 7 students not currently

enrolled. Participants were divided into three conditions relatively evenly across both mediums, with 48 participants in the control, 40 participants in experimental condition 1, and 39 participants in experimental 2 out of the Vassar students, and 25 participants in each condition for prolific. Participants who made only a single group were excluded from the analysis on the basis that they did not fully understand the task or were not attempting to follow the instructions. This only excluded 2 participants, both of which were in the control condition of the prolific participants, leaving 23 students in this condition.

Stimuli

The stimuli created for this experiment were a set of forty-two phrases describing content related to career development. Each description corresponded to the content available on an actual page of the Vassar College career development website in late 2018. All descriptions were shared with several staff members at the CDO to ensure accuracy. Six pages from the original website were not included because they either did not pertain to student users of the site or they were pages that could be described as "miscellaneous" rather than task oriented. An example of this second case would be the "Handouts" section of the site. All the phrases used on the cards were inspired by discussions with staff members at the career development office, who helped to clarify the purpose of different portions of the site. The descriptions used in the experiment are listed in the appendix (Appendix A).

Procedure

Across conditions, participants received very similar instructions except one particular sentence was manipulated to highlight either a particular goal one might want to complete on the site (the two experimental conditions) or no goal (the control). Two different experimental conditions were used to examine how the particular nature of the goal might change the concept map, not just the fact that a goal is highlighted in general. The manipulated sentence is presented below in its 3 forms.

> Control: The sorting of these concepts will be used to structure an actual website containing this content, so keep this in mind as you form groups.
>
> Experimental 1: The sorting of these concepts will be used to structure an actual website containing this content. Students coming to the site are often particularly interested in securing summer internships, so keep this in mind as you form groups.
>
> Experimental 2: The sorting of these concepts will be used to structure an actual website containing this content. Students coming to the site are often particularly interested in figuring out what types of careers they might enjoy, so keep this in mind as you form groups.

After the initial instructions phase, all participants completed the same card sorting task on Optimal Workshop, an online resource for usability studies. A screenshot of the task can be seen below (Figure 2). The cards contained the 42 phrases described earlier in this section, and participants were asked to sort all the cards into between 7-14 groups. A specific range of numbers of groups was suggested to encourage participants to work at a similar level of abstraction as the original website, which divided the 42 pages of content into 11 subsections. However, there was no minimum or maximum number of groupings required to complete the task.

In order to create groups, participants needed to drag cards from the left sidebar into the right part of the screen. Cards could be moved between groups as much or as little as possible, and groups could be repeatedly renamed. However, each participant was required to use all the cards to create their final groupings and name each of these groups in order to finish the task. There was no time limit, but participants took on average 15 minutes and 13 seconds to complete the card sort.

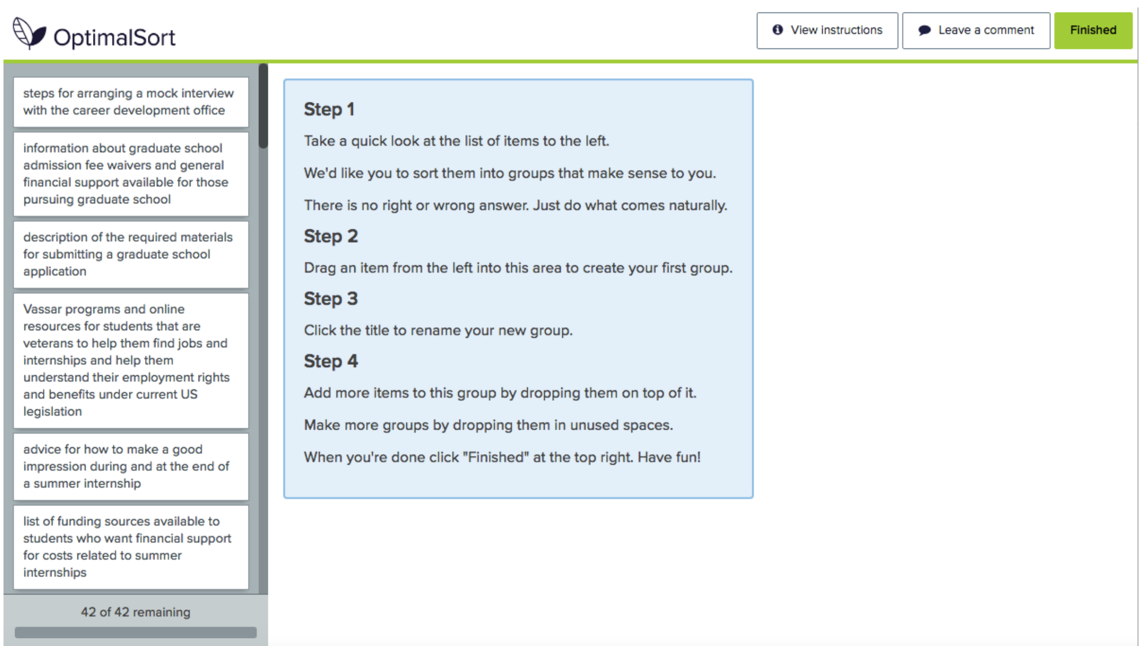

Figure 2. Screenshot of the initial part of the task where some additional instructions were given to help the participant complete the task.

Analysis

In order to evaluate whether the participants were systematically creating different groupings, I used a metric known as “edit distance.” Edit distance is a measure of how different two sets of groupings are from one another, and in particular, it is the minimum number of card moves (edits) that it would take for one set of groups to change into another (Deibel, Anderson, & Anderson, 2005). For example, consider the following two sorts created by person A and person B:

$A_1 = [1,2]$	$B_1 = [1]$
$A_2 = [3]$	$B_2 = [2,3,4]$
$A_3 = [4]$	$B_3 = []$

Sort A can be changed into sort B through at minimum 2 moves. These moves are as follows: take item 2 and move into A_2, and then take item 4 and move into A_2. After these moves, the sorts belonging to person A look like it does below, which is exactly the same as the groupings created by person B.

A_1 = [1]
A_2 = [2,3,4]
A_3 = []

Because this is a measure of difference between a set of sets, this metric allows us to quantify the difference between two sets of groupings. For this experiment in particular, I used edit distance to calculate the average difference between the groups produced in the different conditions. Before this analysis was conducted, I noticed in preliminary evaluations of the data that the Prolific users seemed to create far fewer groups than those made by Vassar students (M = 6.32, SD = 2.44; M = 8.05, SD = 2.39 respectively). A two-sample t-test comparing the number of groups made in each condition confirmed this suspicion ($t(198) = -4.92$, $p < .0001$). For this reason, all comparisons were only made within either the Vassar group or the Prolific group, not across the two student populations.

Given that edit distance can only be calculated between a pair of groupings, I calculated the average edit distance between every pair of conditions within a medium. This meant that for each medium, there were 3 average edit distances, one for every pair of conditions (ex: the pair condition 1 and condition 2 in Vassar students). Each average was calculated by selecting one condition as the reference and calculating the edit distance between each of its groupings and the groupings of the other condition and then averaging over all of these edit distances.

For every average edit distance, i.e. every pair of conditions, (all 6), I constructed a bootstrap sample of average edit distances. This involved randomly assigning each subject's groupings to one of the two conditions and then recalculating the average edit distance with these two randomly created samples. Each bootstrap sample contained 100,000 generated average edit distances.

Results

Between the three Vassar task instruction conditions, there were no significant differences between any pair of groups. The groupings between condition one and condition two had an average edit distance of 18.14 (p = 0.76), the groupings between condition one and three had an average edit distance of 17.22 (p = 0.84), and the groupings between condition two and three had an average edit distance of 16.91 (p = 0.67), but none of these were anywhere close to significance.

Similarly, there were no significant differences between any pair of the three Prolific task instruction conditions. The groupings between condition one and condition two had an average edit distance of 23.71 (p = 0.42), the groupings between condition one and three had an average edit distance of 23.84 (p = 0.52), and the groupings between condition two and three had an average edit distance of 23.95 (p = 0.45), but none of these were anywhere close to significance. Overall, the pattern of results from both mediums suggested that there is no evidence supporting the idea that different tasks contexts may rely on different mental models.

As a follow up to the earlier result suggesting large differences between the groups created by Vassar students and Prolific participants, an average edit distance calculation and bootstrap analysis was conducted between the groups created by these two populations. As expected, the average edit distance between the two mediums was highly significant (M=22.41, $p < 0.001$). This suggests that the individuals in these groups were in fact creating different groupings, which supports my choice to separate them out for analysis.

Discussion

The results of this first experiment indicate that there are no significant differences between the groups constructed under any of the particular task contexts in a given medium. This seems to suggest that regardless of the specific task completed on a site, users have one salient mental model organizing content related to the site's purpose. These results were somewhat different than some of the previous research I mentioned earlier which suggested that mental models might differ quite considerably even in somewhat similar contexts (Ratneshwar, & Shocker, 1991; Ratneshwar, et al., 2001). Why then did my results not corroborate this?

One possible explanation is simply that regardless of the specific task context, users have a salient mental model organizing this information that overrides any requests from the experimenter to envision the organization of such information in a specific context. In the case of the Ratneshwar experiments, participants were asked to rate the similarity of various items under different contexts that, importantly, highlighted vastly different attributes of the objects. For example, in Ratneshwar, & Shocker (1991), participants were asked to name and rate the typicality of snack foods that either might be eaten for breakfast when in a hurry instead of an ordinary breakfast or snack foods that might be eaten at an evening party with friends. In the first case, the key attributes are portability and healthiness, while in the second case, taste is probably the most critical attribute. Under these circumstances, it makes sense that the different contexts evoke different snack foods and different ratings because two snacks can have similar attributes in one area but not in another. In this case, there might be a base organization of snack foods, but these may not be compatible with the different specific contexts due to the different highlighted attributes and thus the contexts result in different organizations.

By contrast, in my experiment the different usage contexts do not necessarily highlight different attributes, but simply increase the salience of certain relationships between concepts. Highlighting relationships does not change the defining attributes used to create the groups, even if it could mean more differentiation within the particular group where these relationships are present. This is different than in the Ratneshwar, & Shocker experiments (1991). In these experiments, the nature of the contexts were such that the concept space might have needed significant morphing to shift from one context to the next, while the contexts in my experiment might not have required much morphing of the space at all. For this reason, a base organization of concepts for career development would not require reorganization. This base organization of concepts still "makes sense" in any of these contexts, so there is no reason to construct anything new for any particular context.

Any changes would be using sub or super groups of the original set of groupings, which would create groupings that are mostly the same. This means that even if users feel that certain relationships are more critical, there might not be any changes to their mental model. This is not necessarily a disadvantage from the intuitive design perspective, as these groupings then reflect an expectation about how the webpage would be organized, even if the fact that participants may see some relationships as being more prominent is not reflected in the groupings.

Another possible explanation is that any changes to the mental model in a given task context were incredibly small shifts compared to the overall variability. In my experiment, there was an incredibly wide variety of mental models between individuals. In the two mediums, the average edit distance between any two groupings regardless of condition was 17.33 for Vassar students and 23.51 for Prolific participants, indicating that there was a quite a large amount of variability even between the groupings made by participants within a medium. If you recall from the earlier discussion of the meaning of edit distance, this indicates that changing one individual's set of groups into another would require moving about 17 and 24 cards on average within the Vassar and Prolific participant pools respectively, which means that close to half the cards need to be moved. This suggests that people varied quite a bit in how exactly they made their groups. To illustrate this point more fully, the titles for two sets of groupings produced by two randomly selected participants from the Vassar student pool are shown below in Table 2. Just the titles are used so that the table is readable.

Table 2.

Examples of groups created by participants.

Groups Made by Participant 1	Groups Made by Participant 2
Interview Techniques (6 items) Finding a Job (5 items) Misc. (1 item) Finding an Internship (3 items) Specific Information for Minority Groups (5 items) Alumni Connections (4 items) Financial Help (4 items) What is the CDO? (2 items) So I've Chosen Grad School (Applying) (5 items) Deciding On Grad School (Where Do I Go?) (3 items) Applying for a Job (2 items) Exploring Future Opportunities (2 items)	Graduate School (5 items) Job Search (15 items) Assistance (5 items) Law School (4 items) Career Development Office (4 items) Vassar Programs (4 items) Career Building (5 items)

As can be seen, these two participants produced vastly different groupings. This is immediately apparent in both the number of groups created and the names of the produced groups. For example, it appears that most of the job and internship related concepts made it into a single group for participant 2, while for participant 1, these concepts were broken into several categories, such as "Finding a Job" and "Applying for a Job." In the case of Law and Graduate schools, participant 2 broke these into two separate groups while it appears that participant 1 treated them as one. Participant 1 even has a group for funding, while participant 2 has no such group. Though these are just two single examples out of the many groupings created in this experiment, they allow us to see a concrete example of the kind of variety we can see when examining the groups. If the systematic changes an individual might make between conditions is small, say only 5 cards or so, it would be completely swallowed up by the between subject differences because I used a between subjects design. I choose this design due to concerns that it would be difficult to get the participant to create groupings under different contexts, both because of suspected lack of motivation from the participants and the difficulty in avoiding dependence between the groupings created, but this means that I could not detect any particularly small systematic changes that would be seen with a within subjects design.

Theoretically, it may be interesting to explore this question further by utilizing a within subjects design that avoids some of these concerns, but this would come with its own set of problems. A study of this sort might be subject to demands effects because the hypothesis of the experiment would be difficult to mask and potentially easy to guess. Regardless, from a practical standpoint, it is probably enough to see that the differences, if any, are quite small compared to the overall variability. It is unlikely that the specific task context affects the mental model in any discernable way given such small differences. Website designers could potentially design sites by asking users specifically to think about critical site behaviors or they could create even dynamic sites that change depending on the task context, but if the differences in the mental models are small under different contexts, it is unlikely that this will make a impact the overall performance enough to warrant a lot of attention.

Overall, the results seem to suggest that there is no reason to suspect that mental models shift in any really noticeable way between task contexts, at least as long as the tasks have non-conflicting goals. In terms of the larger question, this means that designing for a general context related to the website's usage is probably sufficient to be designing for most specific contexts. In this way, what is intuitive to a user in one context is not likely to change drastically as we move to the next, as long as this next context is reasonably similar.

This is also a potentially useful finding for the more general research community interested in mental models. While most of the early research presupposes static mental models, newer research tends to strongly emphasize the dynamic nature of these models instead of what carries over from context to context (Barsalou, 1982; Casasanto, & Lupyan, 2015; Collins & Loftus, 1975). My particular findings highlights the static nature of mental models, potentially because I presented task contexts that while different, were not necessarily in conflict with one another. This could suggest that the dynamic nature of mental models seen in more recent experiments is in part a product of the tremendous differences in the presented contexts. Perhaps when able, people will rely on a more static mental model that acts as a base for the set of items, and only when this base is in conflict with the highlighted characteristics will people make more drastic shifts to the mental model for a given task.

An important detail of the results that we have not yet explored is why there were significant differences in the groupings created by the Prolific and Vassar participants, which was not initially predicted. One possible explanation is that Vassar students may have some previous experience on the career development site that influences their mental models. We would not expect this for the Prolific participants.

This was explored by using sampling with replacement to calculate the average edit distance between sampled participants in each medium and a set of groups representing the structure of the current career development office site, and then taking the difference between these edit distances (Prolific minus Vassar). This was done 100,000 times. The resulting confidence interval indicates that the Vassar students did in fact produce groups that were more similar to the current career development website structure than the groups produced by the Prolific participants (95% CI [5.13, 8.36]). However, a closer inspection of the average edit distances between the groups produced by the Vassar students and the actual career development office website groups indicated that there was no evidence to suggest that these groups were the same (95% CI [17.02, 18.92]), as zero was not in the interval, and instead, this confidence interval seems to indicate very large differences between the two. Specifically, this indicates that close to half the cards need to be moved on average to get from a Vassar student's set of groupings to the real groupings used in the career development office website. In fact, the average difference between the career development office website groups and the groups produced by Vassar students was about the same as the average difference between any of the groupings produced Vassar students. This seems to suggest that there is not very strong evidence to indicate that students were influenced by the real career development office website.

Another plausible, and perhaps more likely, explanation is that Vassar students simply devoted more effort to creating the groups than the Prolific participants and had greater familiarity and connection with the task, which resulted in different groupings. This topic, after all, is closer to home for Vassar students than the subjects used on Prolific, even though these subjects were also students. In fact, Vassar students did make significantly more groups on average than Prolific participants, which seems to indicate looking more carefully at how to differentiate between concepts and thus more dedication to the task. I did note, however, that a Levene's test comparing the response times between the two mediums did not yield significant differences ($F(1,195) = 0.021$, $p = 0.88$). However, because response time did not seem as reliable of a metric here (participants completed the experiment on their own time so they could take as many breaks as they pleased), it seems more critical that the number of groups was so different. Additionally, none of the Vassar students were excluded for making only a single group, while several Prolific participants (2) seemed to be trying to skip through the task by making one large group only.

Vassar students, after all, have far more to gain by using additional effort in this task and are more directly affected by the results of the study. By giving their very best answers, Vassar students might potentially create a better usability experience for themselves and other students in the real world rather than just in this particular experimental setting. The career development office website is also a particularly important resource for these students, as it contains many resources that are helpful for life at Vassar and beyond. While the discussion of the differences between Vassar and Prolific participants is not necessarily relevant to the research question, it is certainly relevant for how I proceed forward in improving the usability of the case website. As I continued, I focused on gathering and using data from Vassar students rather than both groups.

Part 2: Mapping Performance to Concept Map

In this experiment, I sought to explore the relationship between mental model accuracy and performance on a task that might utilize this mental model. I was particularly interested in the exact form of this relationship and whether it would hold for multiple measures of task performance. Based on earlier research suggesting that people often use simplified models to navigate the world (Gigerenzer, Todd, & Group, 2001; Kahneman, 2015; Rensink, 2005), I would not expect a simple linear relationship to hold all the way through the space (if at all). Instead, it might be the case that the relationship between mental model accuracy and performance appears random (people are not heavily relying on their mental model compared to other resources like heuristics) or that some sort of spline model would be appropriate to fit here (being very off hurts a lot but being a little better does not help your performance much). Either of these scenarios might be the case if only a very loose similarity between the mental model and the website makes any positive difference on performance.

Method

Participants

For this experiment, 55 participants were recruited from Vassar College. Only Vassar College students were used due to the concerns about effort from the previous experiment. Again, all participation was voluntary. In total, there were 14 first years, 12 second years, 13 third years, and 16 fourth years. Participants were divided into the 2 conditions described below as evenly as possible with 25 in the first condition, which used the same website structure as the current career development office website, and 30 in the second condition, which used the website structure created using the data from the first experiment. Of these participants, only 2 had used the career development more than 4 times over the course of the previous six months. Based on this, I was not concerned that the participants would be overly influenced by previous experience on the career development office site when they attempted to navigate the website structures presented here.

Two participants were excluded from the analysis. One participant did not complete the second task in the experiment, and another only made a single group, indicating low effort in the concept-mapping task. This left 24 participants in the first condition and 29 in the second. In addition, any individual trials that took over 5 minutes to complete were excluded from any analysis involving time to completion. This was done because trials longer than this seem to indicate that the student stepped away from the task or their computer for some reason rather than that they were still searching for a page.

Stimuli

The first part of this experiment was exactly the same as the control condition in the first experiment. The very same set of descriptions from the first experiment was used for the concept-mapping portion of this experiment, and no changes were made to the instructions (Fig 2).

For the second part of the experiment, 2 website navigation bars were constructed using the Treejack tool on the Optimal Sort website. The two navigation bars used during this experiment are shown below (Figure 3). The first navigation bar is exactly like that of the real career development website office, except that it contains only the headers that are relevant to the chosen list of concepts, and two pages that appear twice in the navigation bar only appear once here. This was done so that the edit distance analysis method could still be used. One of these pages, internship funding, was removed from the funding page so that it would be in nearly the same place in the two different website structures. The hope was that by keeping the page that was in the same place in both websites, removing one of these copies could not hurt the original website's performance in contrast to the new one. The second page, informational interviews, did not have either of its copies in the same place as the new website, so one of the pages was randomly removed.

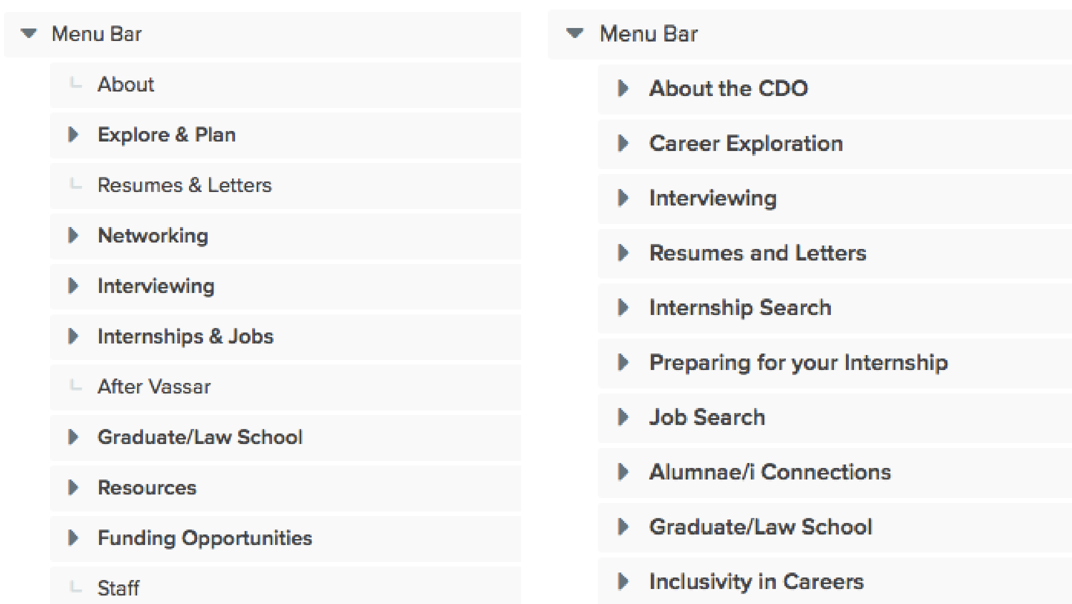

Figure 3. The two navigation bars used in the experiment. The top navigation bar resembles that of the career development office and the second is the one created using Vassar student data.

The second navigation bar was constructed using the data collected from Vassar students in the first experiment. Only data from Vassar students were used due to the suspicion that Vassar students devoted more effort to the task and because Vassar students are the most relevant user group for this website. The website structure was created by taking this data and using complete hierarchical clustering to create aggregate groups. Complete hierarchical clustering was chosen as the clustering method because the groups created by this clustering method tend to be highly compact, thus the individual items in the group tend to be quite similar (Boehmke, 2017; Kilitcioglu, 2018). The groups that exist at the 75% dissimilarity level (the maximum allowed dissimilarity between any two items in the group) were used as the basis for the website structure. These are the groups depicted in the below dendogram (Figure 4). This dendogram provides an illustration of the order by which concepts and groups are added together to form larger groups, with groups that are added together earlier connected at the bottom and groups added later connected at the top.

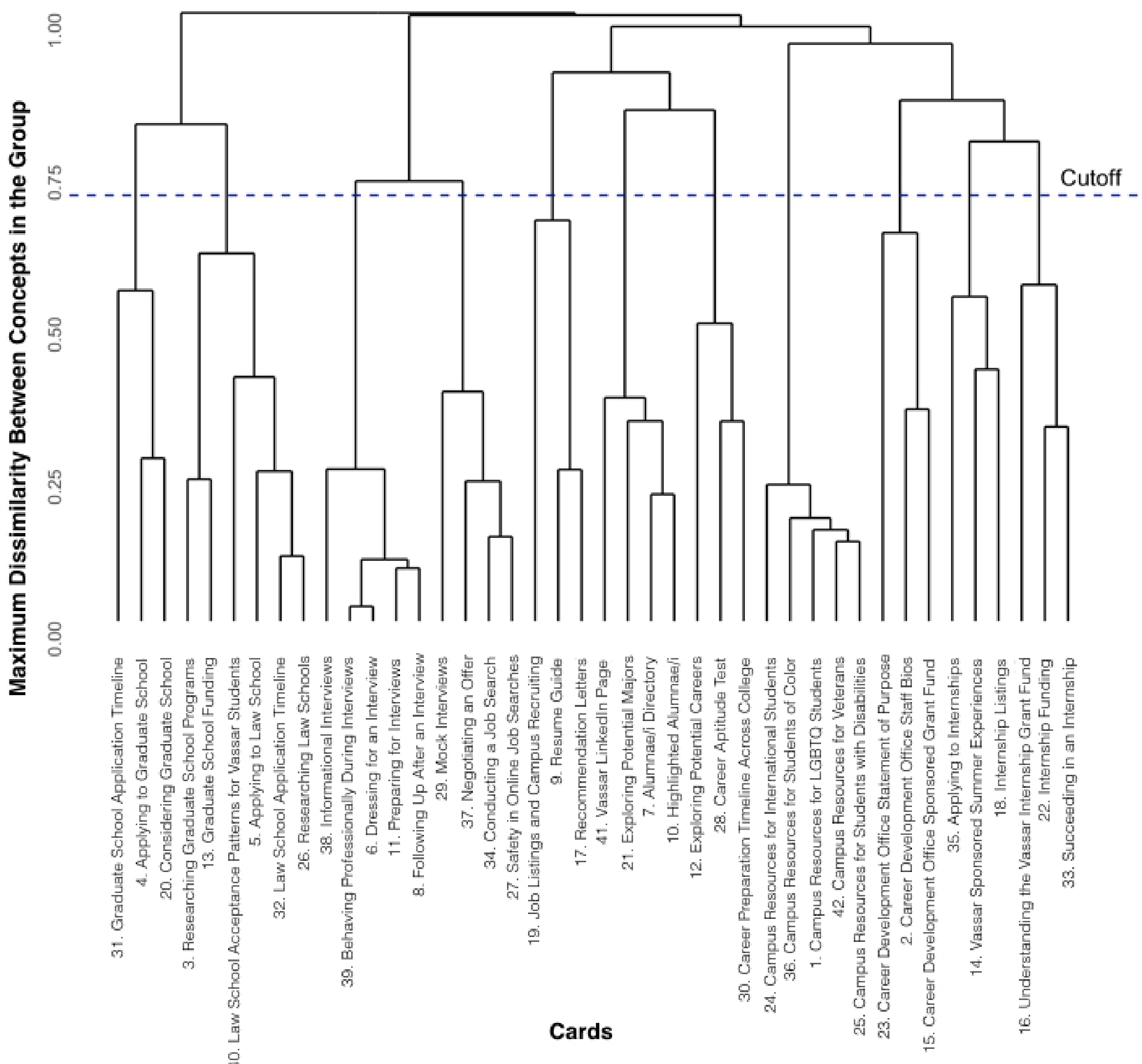

Figure 4. Dendogram depicting a hierarchical structure created from participant data. Each page is labeled with a shorter few word description to help with readability. The complete page descriptions that correspond to each number can be found in the appendix (Appendix A). The cutoff indicates at what level of dissimilarity the original groups were constructed from.

Several changes were made to these groupings to produce the final set used to make the website structure. Firstly, the "Job Search" page was moved from its original place in the "Letters and Resumes" group into the "Job Search" group. This concept was added to the "Letters and Resumes" group very late into the process, close to the 75% dissimilarity mark, which suggests that it was not consistently placed in any group. In addition, several concepts related to graduate school ("Graduate School Funding" and "Researching Graduate Schools") were slightly moved. Initially, the dendogram produced 2 groups, which corresponded, for the most part, to one group of containing graduate school information and one containing law school. However, these two items ended up in the law school group. It seems like this was the case because a number of participants grouped this set of items into groups corresponding to graduate school and law school and a number combined graduate and law school together into one group. This seemed to result in the strange end groupings.

For this reason, all the concepts from law and graduate school were initially combined into one group, and then split into 2 subgroups so that all the graduate school concepts were in one group and all the law school concepts were in another group. This was done so that both kinds of main groupings could be respected.

Names for each of the groups were identified by taking each aggregate grouping and calculating the number of overlapping items between it and groups constructed by participants. The name of any grouping created by participants with over a 60% match was included in the list of possible names. Using this set, names were either directly selected, created by combining several names, or were inspired by some of the answers given. The final set of groupings can be seen in the appendix (Appendix D).

Procedure

Participants began the experiment by completing the same concept-mapping task as described in the first experiment. I used the same set of instructions as the control and did not change any elements of the task from this except that an additional question was added to the beginning of the experiment asking about experience on the career development office site. The second part of the experiment involved navigating through a navigation bar to find various pieces of information. An example of this can be seen above (Figure 3). On each trial, participants could explore as much as much of the navigation bar as they liked and could undo their answer until they locked it in and moved on to the next question.

In total, participants were asked to find 12 different items, but they were not given any feedback on whether they had correctly identified the location of each item. The 12 requested items are listed in Table 3 in their question form. These items were presented to the participants in random order so that the influence of experience navigating through the navigation bar could be controlled for any given item.

Table 3. *List of questions/item given to participants.*
1. Where would you expect to find information about funding meant to help support students doing low and unpaid summer internships? 2. Where would you expect to find information about financial support for graduate school? 3. Where would you expect to find information about funding that comes directly from the CDO for students at any point in their internship or career search?

4. Where would you expect to find information about summer research opportunities at Vassar?
5. Where would you expect to find advice about when in your college career to begin searching for internships?
6. Where would you expect to find information on how to set up a mock interview with the career development office?
7. Where would you expect to find a page describing how to request and submit a letter of recommendation from faculty?
8. Where would you expect to find examples of the careers that students who graduated with your major are doing now?
9. Where would you expect to find a link to a list of all previous Vassar alumnae/i?
10. Where would you expect to find a page listing companies that are actively recruiting people of color?
11. Where would you expect to find a career test tool to help you find careers you might be interested in?
12. Where would you expect to find tips for how to conduct interviews with individuals in your field of interest to learn about their work?

The items referred to in the first nine questions are all items that could either logically appear in multiple places in a given navigation bar or are in different places in the two navigation bars. Out of these questions, all items but item 5 appear in different places in the two navigation bars. All 9 of these items could be in multiple places in the original career development office website navigation bar, and 6 (excluding items 2, 5, 8) have multiple logical locations in the newly created navigation bar.

The final three items refer to items that are in the same place in the two navigation bars, but are in subsections that have different names between the two versions. The purpose of these items was to explore the importance of naming when creating groups, as it is possible that naming is also a critical part of matching the user's concept map.

Analysis

In order to evaluate the relationship between the accuracy of the mental model and the aptitude navigating the site, three values were calculated for each participant. Firstly, the similarity of their mental model to the structure of the site was calculated using edit distance, just in the same way as in the first experiment. That involved taking the participant's groupings and calculating its edit distance with respect to the groupings of the navigation structure used in their condition.

Secondly, each person's overall accuracy in the tasks and the speed of their correct answers were also calculated to be used as measures of aptitude. Accuracy was considered to be correctly identifying the location of a particular piece of web content. Both of these measures were used because they seemed like plausible metrics. Accuracy seemed like the more important metric of the two, but speed of correct answers could also be an important differentiator if the two conditions did not differ on accuracy. This might be the case if the accuracies are both quite high. Initially, a linear model predicting each of these values (accuracy and speed of correct answers) using edit distance was fit. After initial fitting of these models, some exploration of possibly improved models was conducted.

In addition to these measurements, the accuracy and speed of correct answers was also calculated for each task individually and each website condition overall to explore the usability of the two website structures. The analyses conducted with these values were much more exploratory, and only two two-sample t-tests were planned in advance, one comparing the accuracy between participants in the two groups and one comparing their speed of correct answers.

Results

<u>Main Analyses</u>

To explore the relationship between mental model accuracy and site navigation aptitude, the exact nature of the relationship between mental model accuracy and accuracy and speed of correct responses were evaluated. First, a simple linear model was fit to predict accuracy using edit distance for each participant. The results of this regression indicate that edit distance is a significant predictor of accuracy (B = -0.012, $F(1,57)$ = 10.71, p = 0.002). Specifically, edit distance has a negative relationship with accuracy; when edit distances are larger, accuracies are lower. To explore this relationship further, two additional models were fit, one using a spline due to the change in slope observed in the graph at about an edit distance of 14, and one using a quadratic predictor, to fit the potential additional decrease at the far end of edit distance. Each model was fit and their AIC values were compared.

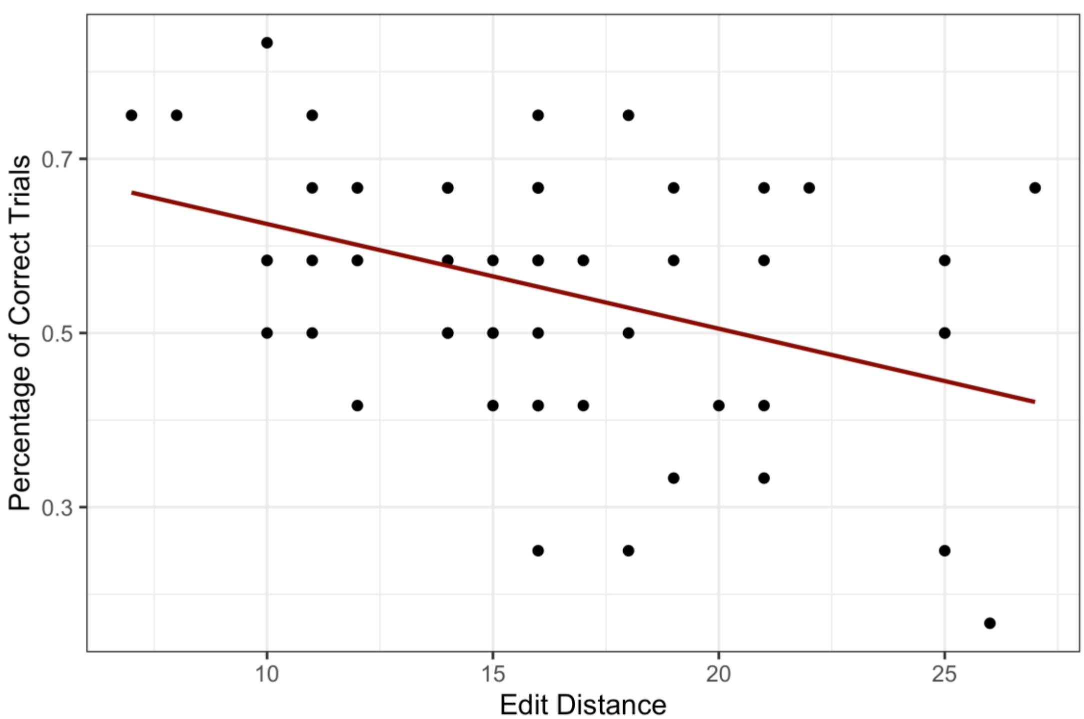

Figure 5. Graph depicting the relationship between edit distance the percentage of correct trials for each individual person. The line for the standard linear

model is graphed here to make the negative relationship more clear.

Neither of the two additional models had an AIC value lower than the original, indicating that they are slightly poorer models, though the AIC values were somewhat similar (Table 4). This suggests that the original simple linear regression model, while being technically the best one though a direct AIC comparison, is really not all that distinguishable from the others.

Table 4.

Summary of model fits predicting percentage of correct responses with edit distance.

Model	F-statistic	P-value	Adjusted R^2	AIC
Simple Linear	F(1,57) = 10.71	0.002	0.14	-69.70
Spline	F(2,56) = 5.48	0.007	0.12	-68.08
Quadratic	F(1, 57) = 9.15	0.004	0.13	-68.33

Another simple linear model was fit to predict response time for correct trials using edit distance for each participant. The results of this regression indicate that edit distance is a significant predictor of correct response time ($F(1,57)=10.71$, $p = 0.006$). Specifically, edit distance has a positive relationship with response time for correct trials; when edit distances are larger, response times are slower. To explore this relationship further, three additional models, one using a quadratic predictor, one using a cubic predictor, and one using a quartic predictor, were fit in order to try to match the appearance of an increasing slope in the second half of the graph. Just like for the previous models, their AIC values were compared to determine the best fit.

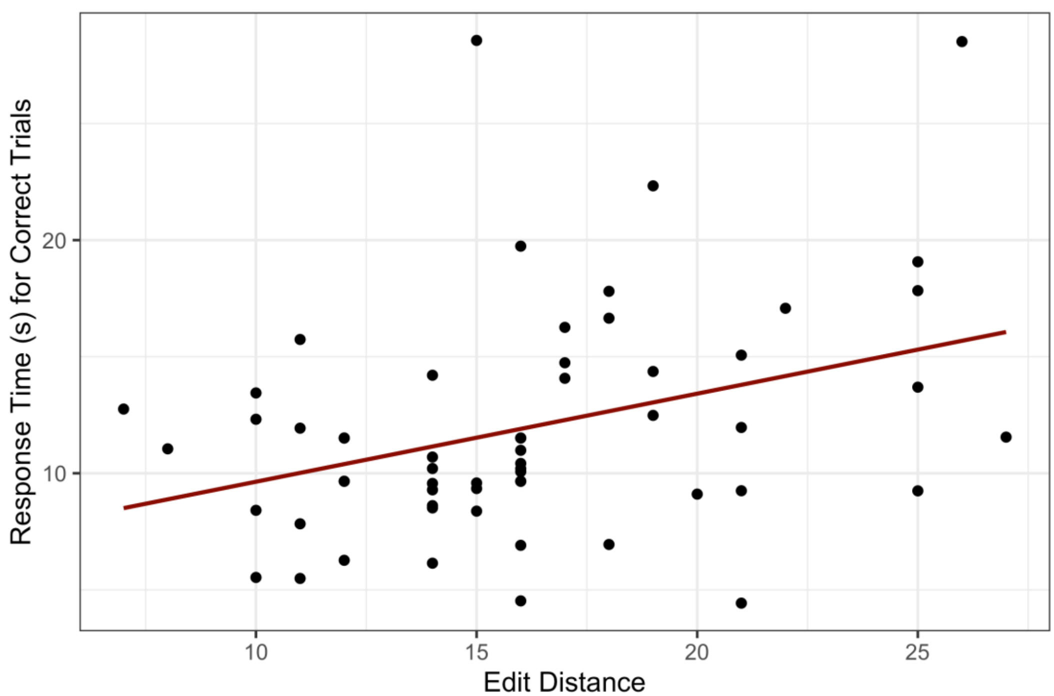

Figure 6. Graph depicting the relationship between edit distance and the response time of correct trials for each individual. The line for the standard linear model is graphed here to make the positive relationship clearer.

Based on the measures of AIC, the cubic model is technically the best predictor of response time for correct trials. However, it is important to note that all the AIC values are quite similar to one another, so there is no real way to distinguish the exact form of the relationship (Table 5).

Table 5.

Summary of model fits predicting percentage of correct responses with edit distance.

Model	**F-statistic**	**P-value**	**Adjusted R^2**	**AIC**
Simple Linear	F(1,57) = 8.14	0.006	0.11	353.24
Quadratic	F(1,57) = 8.89	0.004	0.11	352.56
Cubic	F(1,57) = 9.00	0.004	0.12	352.46
Quartic	F(1,57) = 8.74	0.005	0.12	352.70

The two website structures, the original career development office site and the new site structure, were compared using 2 sample t-tests for both accuracy and speed of correct responses. The two sample Welch's t-test for accuracy revealed that participants were significantly more accurate in the new website structure than they were in the old ($t(44.98) = -5.08$, $p < 0.001$). However, a two sample Welch's t-test for speed of correct answers indicated no significant difference between the groups ($t(37.89) = 1.11$, $p = 0.27$). This indicates that there is no evidence that there is any difference between the two groups in their speed of correct answers.

<u>Secondary Analyses</u>

In the set of tasks, 8 tasks were explicitly designed to have items in different positions in the two website structures. In particular, these were tasks 1 through 9 excluding task 5. Since these were the tasks that primarily highlighted the differences between the two website structures, a t-test comparing the each participant's average performance on these trials for the two website versions (version 1, the original, versus version 2, the new) was conducted. The test revealed that participants in version 2 significantly outperformed those in version 1 ($t(39.129)=4.91$, $p < 0.001$). In fact, in this sample, the participants using the 2nd website structure performed a whopping 19.08% better than those using the 1st website structure (*CI*: [11.22, 26.94]). This seems to suggest that the 2nd website structure placed these items in more intuitive locations overall.

One additional point of interest was whether or not naming seemed especially important for effective navigation. Three of the tasks referred to items in the exact same location in the two website structures, but using different names. The names used in the second website navigation bar were taken directly from the names of similar groups produced by students from the first experiment. To explore the effect of naming, a t-test comparing the each participant's average performance on trials 10 through 12 for the two website versions (version 1, the original or 2, the new) was conducted. The test revealed that participants in version 2 significantly outperformed those in version 1 on these questions ($t(50.8) = 2.30$, $p = 0.026$). In fact, in this sample, the participants using the 2nd website structure performed 15.57% better than those using the 1st website structure (CI: [2.00, 43.06]). This seems to suggest that these new names are potentially more clear and intuitive to the participants than the names originally used by the career development office.

Discussion

The results from this experiment strongly support that there is in fact a relationship between the similarity of a mental model to a website structure and performance navigating that website. In fact, this relationship was seen in regards to two different measures of performance: overall accuracy (can you find what you are looking for) and speed of correct answers (how fast are you when you are right). Both of these relationships reflected a very intuitive pattern: the more alike your mental model is to your world, the better you can navigate this world. While there are some previous results that corroborate this view, these studies do not use quite so detailed a measure of mental model similarity as was used here, instead opting for subjective evaluations or posing simple questions about causality to the participants as their metric of similarity (González, & Juárez, 2013; Qudrat-Ullah, 2014). The method used in this experiment also allowed me to track similarity as a continuous variable so I could see if the relationship changed in any way in different areas of the similarity space. I saw some evidence for a relationship that is different than a simple linear model in the case of response time, but differences between the models was not large enough to be definitive. Thus it is possible that response times increase at a quicker rate as difference increases, but it is also quite possible that it does not. For accuracy, I also could not completely distinguish between the forms of the models. All I can say is that there does seem to be a positive relationship between edit distance and response time, and a negative relationship between edit distance and accuracy. This does, however, suggest that the more intuitive website structures are more similar to the mental models of their users, even if the exact form of the relationship is not known.

Notably, all of the models had somewhat low r-squared values, indicating that the models did not fit the data very closely. This has an important practical implication, which is that creating a website structure that matches closely with users' mental models does not ensure that they will have ease navigating the site. In particular decisions like the placement of items, fonts, and colors are all likely to substantially affect performance, particularly because individuals do not have the same preferences or expectations across cultures or demographic groups (Barber & Badre, 2018; Reinecke & Gajos, 2014; Mead, Spalding, Sit, Meyer & Walker, 1997). This means, not surprisingly, that you are not done improving your website even when the website structure is highly similar to your users' mental models.

Importantly, I also saw support for evidence that the card sorting task does in fact produce more intuitive website structures. Participants using the second navigation structure were more accurate than those who used the original, which supports the view that a student created navigation structure is more intuitive to students. This was additionally supported by my two follow-up analyses.

The first of these showed that individuals using the new navigation structure were more likely to find the correct page than individuals using the old structure for all the trials differentiating the two structures. This suggests for the trials that really mattered, performance was better in the new structure.

However, not all individual tasks had higher accuracies for the second website structure than the first in this sample. In particular, in this sample individuals using the original website performed better than those using the new one for task 3 (CDO office funding) and task 6 (mock interviews). Thus it is certainly possible that while the overall effect is in favor of the new website structure, a few of the pages are not in their optimal locations. One possible explanation for this is a particular constraint of the chosen method I used to do hierarchical clustering. The complete method is performed using pairwise comparisons of individual items, rather than groups or subgroups, so it is possible to construct aggregate groups that were not made by any individual person (Boehmke, n.d.; Kilitcioglu, 2018). This means that while whole groups may seem mostly coherent, certain items may not fit the overall pattern.

The second of my follow up tests also supported the success of the newly constructed navigation structure. This test demonstrated that participants using the second navigation structure were still more likely to find the correct page than participants in the first navigation structure when the actual location of the page was exactly the same but the submenu had a different name. In the case of these tasks, all three had the same or higher accuracies in this sample using the 2nd website structure.

The names of the submenus were created at best directly and at worst indirectly from the names of similar groupings provided by students. This suggests the intuitive design process is not just about creating a better website structure in the traditional sense, but making the structure more apparent by providing appropriate names. This particular finding is very important in the context of this study because the naming of groups is not given the attention that it perhaps deserves in many concept mapping tasks. Some methods of concept mapping, such as calculating word co-occurrence or certain kinds of interviews, do not explicitly ask the user how they would name groups at all (Gonzalez & Juarez, 2013; Goodyear et al. 2005). Even in the case of card sorting, asking users to name their groups is not required in all circles (Goodman et al. 2012).

In a broader sense, this result suggests that the category names are an incredibly important part of understanding the underlying mental models that help us navigate through the world. This is important because research on mental models is often focused on the kinds of relationship between objects, such as what features they have in common and how important these features are in a given context, rather than how the particular names of categories encapsulate or leave out these relationships (Barsalou, 1982; Johnson-Laird, 1980; Ratneshwar, et al., 2001; Ratneshwar, & Shocker, 1991). This may mean that researchers are not devoting enough in these studies to understanding what words participants believe tie categories together. Names of categories, just like the concepts they contain, invoke a variety of associations that can affect what kinds of concepts we imagine to be contained within them.

My other metric of website performance, however, did not illustrate the same pattern described above. There was no difference in the response time on correct trials for participants in either group, indicating that individuals performed the same on this metric regardless of the website structure used. One possible explanation for this is that response time is affected by many additional factors other than just expectations about the locations of items in the navigation bar. For example the depth by which these items are nested has a tremendous impact on response time. This is the case simply because response times were quite small and opening submenus takes time. For example, in the second website structure, there is a submenu called “About the CDO,” while in the first website structure, most of the items present here are visible at the outer most level of the structure. This means that exploring concepts under this topic takes an additional step in the second website structure compared to the first.

Response times are also strongly affected by outside distracting factors, particularly because participants completed this study on their own time in any location they might like. This was important in order to get a large enough sample, but it does mean that participants could have frequently taken breaks from the experiment. A close look at the data shows that there were certainly trials in the minute to several minutes range, which while reasonable, does seem slightly long given the ease of the task. In may be that participants were not worried about doing the task quickly. In fact, some distractions that do not seem like they would affect accuracy, such as eating lunch while doing the experiment, would strongly affect response time.

Regardless, the positive result for accuracy seems to be strong support for a more usable site. It is far more important that people find things more often than that they do so more quickly. Thus the card sorting method did produce a site that was easier to use for students than the site that existed prior, which indicates that the method is relatively sound. Moreover, it supports the postulated belief from the introduction that performing the method in this way, i.e. with more emphasis on sound scientific practices, can result in improved websites and help us understand some of the causal factors behind the improvement.

Conclusion

The purpose of this thesis was to try to set out and answer the question “what makes a website intuitive?” In order to do this, I reviewed how this question is normally approached in industry and some areas where this current strategy could be improved with influence from academia. I showed that affordance theory provides a strong theoretical framework for helping us ask more specific questions about intuitive website designs in a particular context, and that scientific practices may help us make more sense of the research we do to understand how users use particular products. I also illustrated how a critical evaluation of these practices can help us understand some of the underlying assumptions of the method. I then took a particular case study, the Vassar career development office, and evaluated it with some questions related to affordance theory and proposed a method for making it better, card sorting, with some additional influence from the scientific method.

My goal was to make this website more intuitive, which is the end goal of most research studies related to the user experience. But my goal was also to understand *why* it was more intuitive to the user. In particular, I asked two specific theoretical questions that lie at the base of the card sorting method. Each question could help us identify what features really matter for building conceptual alignment between a user and a website. First, I asked whether a user’s underlying mental model is highly malleable or somewhat static under various task contexts. This was an important starting point to understand conceptual alignment because it helps us see if we need to design for particular situations or if we can design for more general domains. Without this knowledge, it would not be clear how to interpret measures of performance that require participants to complete multiple tasks on the site. I then asked if a website structure was more intuitive precisely because it was highly similar to the user's own mental model, or if the relationship between performance and mental model accuracy is more complicated than this. This was an important question because it is a fundamental assumption in nearly all studies designed to understand the user: the more we match their conceptualization of the site, the better they'll do.

I found that users did not alter their mental model significantly between task contexts. This suggests that when participants are performing a task such as website navigation, they do not need to make any significant shifts in their mental model even when their goals change. This is important because it suggests that we can design websites for certain domains of concepts rather than for specific tasks. I also found, however, that there was a tremendous amount of individual variability in the mental models produced. This suggests that even if we do not have to vary our website structure by task, we will have to try to account for the fact that what is intuitive design for one person may not be for another. While aggregate groups do reflect patterns supported by many participants, this does not mean that all or most participants will necessarily perfectly match this new structure.

I also found that there was indeed a linear relationship between mental model accuracy and performance, and this pattern held for two different measures of performance: accuracy and speed of correct responses. I explored the possible forms of this relationship, but unfortunately, I was not able to find one model that was definitively better than the others. A fruitful point of future exploration would be to spend additional time planning for an experiment that could really differentiate between the possible models. This might involve collecting more data or spending more time understanding what form of the model is really motivated by past experimental results in the field.

Just as importantly, I found that I was able to create an improved website structure using the card sorting method. The new website structure can be seen in my OSF repository at the following link: https://osf.io/sd5hw/. Participants could more accurately identify the location of information in the new site than in the old, and the differences seemed large enough that they were not only significant, but meaningful. The difference in actual structure and the difference in naming both led to significant increases in performance on the site corresponding from about 15% increases up to 19% increases in overall accuracy. Thus my case study was successfully able to illustrate my proposed process from start to finish. I was able to identify a problem using affordance theory as a framework, propose a method inspired by combining industry and academia, and improve the website while also answering deeper questions about why it was improved.

Often the differences between applications and research are stressed significantly more than their similarities. In preparing to write this thesis, I stumbled upon many articles, both scientific ones and blog posts, emphasizing the contrast between these two fields, particularly in their goals, or articles highlighting the barriers that these two fields have to overcome to work together (Kirchherr, 2018; Mallonee, Fowler, & Istre 2006; Schutlz, 2016; Wood & Wood, 2008). The purpose of science, these authors say, is to get at some objective truth, with much less emphasis on the practical applications. By contrast, the authors believed that in user research, the purpose is fundamentally to determine how a product should be designed, with much less emphasis on objectivity and truth. These conflicting goals mean that when push comes to shove, the scientist is content if his knowledge never leaves the lab and the user research is content to make things that work but are not fully understood.

There is no denying that there is some truth in these concerns. But successful collaboration is possible, and there is more overlap in their goals than it initially seems. As illustrated in this thesis, the two can work together quite well when their respective strengths and weaknesses are respected. User experience research can benefit from academic resources that provide guiding research questions and theoretical frameworks relevant to exploring the user's needs and interests. In this way, user research can utilize science's desire for objectivity very directly; the theories proposed by scientists are meant to be domain general and thus applicable in a wide array of settings. Conversely, academic research can also benefit from user-experience studies that further test the validity of the theoretical frameworks emerging out of academia in more realistic contexts. These extensions will demonstrate whether the theories hold water outside of highly controlled lab settings. In this way, academia can benefit from user research's emphasis on testing in real-world situations.

Thus while initially it may seem like the goals of these two fields are contradictory, they are in fact quite complementary. After all, user research and academic research are fundamentally interested in the same thing: gaining a better understanding of people. Collaborate research efforts, such as what has been described here, will only help us deepen and complicate our knowledge of human behavior.

References

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84*(3), 261-271.

Andrews, J. K., de Leeuw, J. R., Larson, C., & Xu, Xiaoqing. (2017). A preliminary p-curve analysis of learned categorical perception research. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*. 1550-155.

Bærentsen, K. B., & Trettvik, J. (2002). An activity theory approach to affordance. *Proceedings of the Second Nordic Conference on Human-computer Interaction.*

Barber, W. & Badre, A. (1998). Culturability: The merging of culture and usability. *Proceedings of the 4th Annual Conference on Human Factors and the Web.*

Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts. *Memory & Cognition, 10*(1), 82-93.

Barsalou, L. W. (2002). Being there conceptually: Simulating categories in preparation for situated action. In N.L Stein, P.J. Bauer & M. Rabinowitz (Eds.), *Representation, memory, and development: Essays in honor of Jean Mandler*. 1-15, Mahwah, NJ: Erlbaum.

Bernard, M (2001). Developing schemas for the location of common web objects. *Usability News 3*(1).

Bernard, M., Lida, B., Riley, S., & Janzen, K. (2002). A comparison of popular online fonts: Which size and type is best? *Usability News, 4*(1).

Beuttner, K. (2018). Personal Interview.

Boehmke, B. (2017, March 03). Hierarchical Cluster Analysis. Retrieved from https://uc-r.github.io/hc_clustering

Boroditsky, L. (2000). Metaphoric structuring: Understanding time through spatial metaphors. *Cognition ,75*(1), 1-28.

Capra, M. (2006). *Usability Problem Description and the Evaluator Effect in Usability Testing* [Doctoral dissertation]. Retrieved from Virginia Polytechnic Institute.

Casasanto, D. & Lupyan, G. (2015). All concepts are ad hoc concepts. In The Conceptual Mind: New directions in the study of concepts. E. Margolis & S. Laurence (Eds.) 543-566. Cambridge: MIT Press.

Chaparro, B., & Bernard, M. (1999). Sitemap design: Alphabetical or categorical? *Usability News 1*(2).

Chen, Y. (2018, June 07). UX Design in the Agile Environment. Retrieved from https://blog.prototypr.io/design-in-agile-2204c5943580

Coltheart, V. (1999). Introduction: Perceiving and remembering brief visual stimuli. In V. Coltheart (Ed.), Fleeting memories (pp. 1–12). London, England: MIT.

Conceição, S. C., Samuel, A., & Biniecki, S. M. (2017). Using concept mapping as a tool for conducting research: An analysis of three approaches. *Cogent Social Sciences, 3*(1).

Deibel, K., Anderson, R., & Anderson, R. (2005). Using edit distance to analyze card sorts. *Expert Systems,22*(3), 129-138.

Eerland, A., Guadalupe, T. M., & Zwaan, R. A. (2011). Leaning to the left makes the Eiffel tower seem smaller: Posture-modulated estimation. *Psychological Science, 22*(12), 1511-1514.

Gaver, W., (1991). Technology affordances. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 79-84.

Gibson, J., J. (1979). The Theory of Affordances In *The Ecological Approach to Visual Perception* (pp. 127-137). Boston, MA: Houghton Mifflin.

Gigerenzer, G., Todd, P. M., & ABC Research Group. (2001). *Simple heuristics that make us smart* [Abstract]. New York: Oxford University Press.

González, V. M., & Juárez, R. (2013). How do you understand twitter?: Analyzing mental models, Understanding and learning about complex interactive systems. *Human Computer Interaction Lecture Notes in Computer Science,* 103-110.

Goodyear, R. K., Tracey, T. J., Claiborn, C. D., Lichtenberg, J. W., & Wampold, B. E. (2005). Ideographic concept mapping in counseling psychology research: Conceptual overview, methodology, and an illustration. *Journal of Counseling Psychology, 52*(2), 236-242.

Goodman, E., Kuniavsky, M., & Moed, A. (2012). *Observing the user experience: A practitioners guide for user research*. Amsterdam: Morgan Kaufmann/Elsevier.

Hartson, R. (2003). Cognitive, physical, sensory, and functional affordances in interaction design. *Behavior and Information Technology, 22*(5). 315-338.

Hawkins, J., & Blakeslee, S. (2008). On intelligence: How a new understanding of the brain will lead to the creation of truly intelligent machines. New York: Times Books/Henry Holt.

Hornecker, E. (2005). A design theme for tangible interaction: Embodied facilitation. *Proceedings of the 9th European Conference on Computer Supported Cooperative Work,* 23-43.

Hudson, H. (2018). Personal Interview.

Johnson-Laird, P. (1980), Mental Models in Cognitive Science. *Cognitive Science*, *4.* 71-115.

Kahneman, D. (2015). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Kaptelinin, V., & Nardi, B. (2012). Activity theory in HCI: Fundamentals and reflections. *Proceedings of the 30th Annual ACM Conference on Human Factors in Computing Systems.*

Kilitcioglu, D. (2018, October 26). Hierarchical Clustering and its Applications. Retrieved from https://towardsdatascience.com/hierarchical-clustering-and-its-applications-41c1ad4441a6

Kim, R. S., Seitz, A. R., & Shams, L. (2008). Benefits of Stimulus Congruency for Multisensory Facilitation of Visual Learning. *PLoS ONE, 3*(1).

Kirchherr, J. (2018, August 09). A PhD should be about improving society, not chasing academic kudos. Retrieved from https://www.theguardian.com/higher-education-network/2018/aug/09/a-phd-should-be-about-improving-society-not-chasing-academic-kudos

Lida, B., & Hull S., Pilcher, K. (2003). Breadcrumb navigation: An exploratory study of usage. *Usability News, 5*(1).

Linde, J. A., Jeffery, R. W., Finch, E. A., Ng, D. M., & Rothman, A. J. (2004). Are Unrealistic Weight Loss Goals Associated with Outcomes for Overweight Women? *Obesity Research, 12*(3), 569-576.

Lindgaard, G., Fernandes, G., Dudek, C., & Brown, J. (2006). Attention web designers: You have 50 milliseconds to make a good first impression! *Behaviour & Information Technology, 25*(2), 115-126.

MacLeod, C., M. (1991). Half a century of research on the Stroop Effect: An integrative review. *Psychological Bulletin, 109(*2*),* 163-203.

Mallonee, S., Fowler, C., & Istre, G. R. (2006, December). Bridging the gap between research and practice: A continuing challenge. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2564411/

McGrenere, J. & Ho, W. (2000). Affordances: Clarifying and evolving a concept. *Proceedings of Graphics Interface 2000*, pp. 179-186.

Mead, S., Spaulding, A., Sit, R., Meyer, B., & Walker, N. (1997). Effects of age and training on World Wide Web navigation strategies. *Proceedings of the 1997 Human Factors and Ergonomics Society*, pp. 152-156.

Nielsen, J. (2017). A 100-year view of user experience. Retrieved from https://www.nngroup.com/articles/100-years-ux/

Norman D. A. (1999), Affordances, conventions, and design. *Interactions, 6*(3) 38-43.

Norman, D. A. (2013). *The design of everyday things*. New York, NY: Basic Books.

Petrie, H., Power, C., Cairns, P., & Seneler, C. (2011). Using card sorts for understanding website information architectures: Technological, methodological and cultural Issues. *Human-Computer Interaction – INTERACT 2011 Lecture Notes in Computer Science,* 309-322.

Piernik, M. (2017, July 14). The major problem with User Experience Research. Retrieved from https://uxplanet.org/the-major-problem-with-user-experience-research-72f0b63716b0

Qudrat-Ullah, H. (2014). Decision strategy and performance in dynamic tasks: The 2-layer process model. *Understanding Complex Systems Better Decision Making in Complex, Dynamic Tasks,* 137-144.

Ratneshwar, S., Barsalou, L. W., Pechmann, C., & Moore, M. (2001). Goal-derived categories: The role of personal and situational goals in category representations. *Journal of Consumer Psychology, 10*(3), 147-157.

Ratneshwar, S., & Shocker, A. D. (1991). Substitution in use and the role of usage context in product category structures. *Journal of Marketing Research, 28*(3), 281.

Reason, J. (1995). Understanding adverse events: Human factors. *Quality and Safety in Health Care, 4*(2), 80-89.

Reinecke, K., & Gajos, K. Z. (2014). Quantifying visual preferences around the world. *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems - CHI 14*.

Rensink, R. (2005). Change blindness. In *McGraw-Hill Yearbook of Science & Technology* (pp. 44-46). New York, NY: McGraw-Hill.

10 Top Principles of Effective Web Design. (2015, February 05). Retrieved from https://shortiedesigns.com/2014/03/10-top-principles-effective-web-design/

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Judgment under Uncertainty,* 3-20.

Schutz, P. (2016, May 23). Spreading the word: Science isn't just for scientists anymore. [Web log post]. Retrieved from https://www.psychologytoday.com/blog/psyched/201605/spreading-the-word

Simmons, J. P., & Simonsohn, U. (2017). Power Posing: P-Curving the Evidence. *Psychological Science, 28*(5), 687-693.

Simons, D. J. (2000). Current approaches to change blindness. *Visual Cognition,7*(1-3), 1-15.

Smith, M., & Loschky, L. (2018). Scene Gist Narrative Priming: Sequential expectations influence scene gist recognition performance. *Journal of Vision,18*(10), 144.

Staggers, N., & Norcio, A. (1993). Mental models: Concepts for human-computer interaction research. *Int J. Man-Machine Studies*. *38*, 587-605.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643-662.

UserTesting. (2017). 2017 UX and user research industry survey report. Retrieved from https://usertesting.com

Vlahovic, T. (2018). Personal Interview.

Wilson, A. D., & Golonka, S. (2013). Embodied cognition is not what you think it is. *Frontiers in Psychology, 4*.

Wood, J. R, & Wood L. E. (2008). Card sorting: Current practices and beyond. *Journal of Usability Studies, 4*(1), 1-6.

Appendix A

List of Concepts Used in the Card Sorting task and Their ID Numbers

This list of concepts was used both in the first experiment and the second experiment. ID is only used during the process of creating aggregate groupings for use in experiment 2.

Name	#
campus resources and general online resources for LGBTQ students related to finding jobs and internships in companies with inclusive work environments	1
career development staff information such as name, email, and title	2
databases and rankings of graduate school programs and suggestions for what to consider when selecting schools	3
description of the required materials for submitting a graduate school application	4
description of the required materials for submitting a law school application	5
descriptions of appropriate attire to wear for an interview	6
directory of Vassar alumnae/i	7
examples and advice for writing thank notes to the interviewer after an interview	8
examples and tips for creating a resume and cover letters	9
highlighted successful alumnae/i and databases of alumnae/i	10
general tips for preparing for interviews and information on the different types of interviews, such as screening interviews, behavioral interviews, or technical interviews	11
guides containing information about and databases of career options	12
information about graduate school admission fee waivers and general financial support available for those pursuing graduate school	13
information about summer work, apprenticeship, and research opportunities at Vassar College	14
information about the grants available to help offset costs for career preparation, jobs and internship search, and graduate and law applications that are provided directly by the career development office	15
information on the internship grant fund, a grant fund sponsored by Vassar College, which provides funding for students taking low or no pay summer internships	16
instructions for how to request and then submit a letter of recommendation	17
internship listings and networking databases	18
job listings and information about off and on campus recruiting	19
key considerations for students thinking about going to graduate school	20
links to Vassar academic department websites, online guides, and the alumnae/i directory to help students connect their major to careers	21
list of funding sources available to students who want financial support for	22

costs related to summer internships	
major goals and services of the career development office	23
online resources for international students, such as information about visa status and employment authorization, lists of fellowships and grants that international students are eligible for, and general advice for jobs and internships	24
online resources intended to help students with disabilities find jobs and internships, secure accommodation in the workplace, and understand their rights under the ADA and related legislation	25
online resources for finding law schools that are a good fit for a student's career interests and academic standing	26
rules of thumb for determining if a job posting found online is legitimate	27
self-assessment tool that evaluates work interests, personality, and skills and then generates a list of careers	28
steps for arranging a mock interview with the career development office	29
suggested timeline for career exploration and preparation across the four years of college	30
suggested timeline for preparing to apply and applying to graduate school	31
suggested timeline for preparing to apply and applying to law school	32
advice for how to make a good impression during and at the end of a summer internship	33
tips for how to be the most successful in a job search	34
suggestions for how to prepare to apply, apply, and interview for a summer internship position	35
support networks and other online resources for students of color to help them prepare for careers and find companies committed to recruiting people of color	36
tips for evaluating and negotiating a job offer	37
tips for organizing and preparing for informational interviews, which are interviews with professionals where the purpose is to learn about their career path	38
tips for professional conduct during interviews and example interview questions	39
top law programs that have accepted Vassar students, common law schools for Vassar students, and information on the law school acceptance rate for Vassar students	40
Vassar LinkedIn page	41
Vassar programs and online resources for students that are veterans to help them find jobs and internships and understand their employment rights and benefits under current US legislation	42

Appendix B

Questions Presented to Staff and Students During Interviews

Questions for Staff:

1. How would you describe the current organization of the website?
2. What do you think students should be looking for on the career development office website?
3. What do you think students are actually looking for?
4. Do you see any mismatches in what information is available on the site and what students tend to look for or what students tend to come to the office and ask about?
5. Other than students, who would you describe as key users of the site?
6. What parts of the website do you think could be improved? What parts do you think are good?
7. Let's imagine we were redesigning the website (unlimited resources) just based on your feedback. What would this site look like? How would information be organized and how would the site look different than it does now?

Questions for Students:

1. What is your major and/or minor?
2. What information do you think should be on the CDO website?
3. What information are you most often looking for when you use the CDO website?
4. (When in your time at Vassar did you look for it)
5. What information have you been able to find using the CDO website?
6. What information have you found difficult to find using the CDO website?
7. Let's imagine we were redesigning the website (unlimited resources) just based on your feedback and needs. What would this site look like? How would information be organized and what information would you focus on?

Appendix C

Lists of Resources Described by Staff Members and Students Related to the Career Development Office Website.

This particular section of the appendix contains notes from all of the interviews conducted during the early part of this thesis. These notes contain particular resources mentioned by either career development office staff or students. Asterisks indicate that an additional student mentioned this resource, and multiple asterisks correspond the number of additional students who mentioned this resource after the first. Some resources are listed in a way that is slightly nested within other resources. The reason for this is that certain students mentioned very broad resources, such as "things for finding a job," while others worked at a much lower level of specificity, saying things like "a list of jobs you could apply for." In order to preserve this in the data, multiple levels of specificity are included.

Staff Responses[1]

Mentioned key resources:

- Upcoming Events (on right side)
- Featured Resources (on right side)
- Resume Guide
- Going Global*
- Announcements
- Grad School info**
- Funding resources/Summer funding*****
- Handshake****
- Interviewing handouts*
- Career Exploration/Explore and Plan*
- Career Prep
- Networking: Alumni + LinkedIn*
- Vault campus(finance/consulting)
- Focus 2

What they think students look for:

- Handshake (may not get to through website)/Internships and Jobs****
- Resume Guide*****
- Funding*
- Networking tab
- Cover letter guides*
- Scholarships
- Fellowships
- Interviewing info (phone and video)

Student Responses[1]

Information that it would be helpful to find on the site:

- Finding job Related/Internships
    - Different jobs you could apply for, ones specific to Vassar*******
        - Include internships from indeed, linkedIn, not just Vass + Liberal Arts consortium*
        - Ones in local area to do during school
        - Information about jobs
            - Job description*, hours, location***, commute time
            - How hands on is it
        - Where to start
    - Internships match interests/major*****
    - Single application for internships
    - Statements from students about specific internships they have done
    - Information on on-campus jobs*
    - Information on how to reach out to get internships/jobs*
    - Alumn contacts for internships
- Finding types of internships and jobs that might be interesting
    - Specific examples of what students do with specific majors****
    - Connections to alumni who have done what interested in****
        - Alumni who are definitely willing to help
        - Link to alumni directory
    - Information about speakers who are coming*
    - Comparing different career paths/jobs*
    - Contacts for professors or admin who might able to help students*
- CDO Office Information
    - Location, numbers**, emails*, hours***
    - What they do***
    - Different types of appointments you can make
    - Staff member info***
    - Times for the CDO events***
        - Be able to register for events online*
    - Make appointments online**
- Prepping for applying
    - Resume Help******
        - Example resumes
        - Industry specific
    - Cover letter help**
    - Interview help****
        - Technical Interviews*

[1] Stars indicate the number of additional individuals who mentioned these resource after the first person

        - Advice from professors on how to do
    - examples from real students for how they prepared for applying
- Specific program information for continuing education***
    - Premed stuff
    - National and international places
- Scholarships/Fellowships***
- Different abroad programs that are related to interests
- Information on research, post bac stuff
    - URSI

What students have used the CDO website for:

- Info on writing cover letter* + resume***
- Looking for internships**
- How to make an appointment*
- LinkedIn/networking stuff
- Interview info

Appendix D

Final Groupings Used to Create the New Website Structure.

Group 1: Interviewing

tips for professional conduct during interviews and example interview questions
descriptions of appropriate attire to wear for an interview
general tips for preparing for interviews and information on the different types of interviews, such as screening interviews, behavioral interviews, or technical interviews
examples and advice for writing thank notes to the interviewer after an interview
tips for organizing and preparing for informational interviews, which are interviews with professionals where the purpose is to learn about their career path

Group 2: Job Search

steps for arranging a mock interview with the career development office
tips for evaluating and negotiating a job offer
tips for how to be the most successful in a job search
job listings and information about off and on campus recruiting
rules of thumb for determining if a job posting found online is legitimate

Group 3: Career Exploration

guides containing information about and databases of career options
self-assessment tool that evaluates work interests, personality, and skills and then generates a list of careers
suggested timeline for career exploration and preparation across the four years of college

Group 4: Alumnae/i Connections

links to Vassar academic department websites, online guides, and the alumnae/i directory to help students connect their major to careers
directory of Vassar alumnae/i
highlighted successful alumnae/i and databases of alumnae/i
Vassar LinkedIn page

Group 5: About the CDO

major goals and services of the career development office

career development staff information such as name, email, and title
information about the grants available to help offset costs for career preparation, jobs and internship search, and graduate and law applications that are provided directly by the career development office

Group 6: Preparing for your Internship

information on the internship grant fund, a grant fund sponsored by Vassar College, which provides funding for students taking low or no pay summer internships
list of funding sources available to students who want financial support for costs related to summer internships
advice for how to make a good impression during and at the end of a summer internship

Group 7: Internship Search

suggestions for how to prepare to apply, apply, and interview for a summer internship position
information about summer work, apprenticeship, and research opportunities at Vassar College
internship listings and networking databases

Group 8: Resumes and Letters

examples and tips for creating a resume and cover letters
instructions for how to request and then submit a letter of recommendation

Group 9/10 (Subgroups of: Graduate/Law School)

Group 9: Graduate School

suggested timeline for preparing to apply and applying to graduate school
description of the required materials for submitting a graduate school application
key considerations for students thinking about going to graduate school
databases and rankings of graduate school programs and suggestions for what to consider when selecting schools
information about graduate school admission fee waivers and general financial support available for those pursuing graduate school

Group 10: Law School

description of the required materials for submitting a law school application

suggested timeline for preparing to apply and applying to law school
online resources for finding law schools that are a good fit for a student's career interests and academic standing
top law programs that have accepted Vassar students, common law schools for Vassar students, and information on the law school acceptance rate for Vassar students