

Effects of Prototype Medium on User Feedback in Product Evaluation

A thesis submitted by

Andrew Stockton Bennett

in partial fulfillment of the requirements for the degree of

Master of Science

in

Human Factors

Tufts University

February 2015

Adviser: Daniel Hannon

Abstract

Understanding User feedback on design concepts can help designers tailor their process more to successfully fulfill user needs and save time and money in the product development process. This paper explores how users understand and evaluate physical prototypes and rendered prototypes of archetypal products across several measures of design quality. Sixteen participants evaluated four products represented with rendered and physical prototypes. It was found that physical prototypes were considered significantly ($p < .001$) more effective at answering design questions across products and attributes.

Table of Contents

List of Tables	v
List of Figures	vi
Effects of Prototype Medium on User Feedback in Product Evaluation	1
INTRODUCTION	2
LITERATURE REVIEW	6
Overview	6
Purpose of Prototypes	6
Influence Of Prototype Fidelity On Design Feedback	9
Influence of Design Method in the Product Evaluation Process	10
Computer Aided Design (CAD)	10
Renderings	12
Physical Prototypes	13
Prototype Representation Medium	13
Design Evaluation	17
Research Gap	18
METHODS	19
Overview	19
Preliminary Studies: Product Selection	20
Preliminary Study 1: Familiar Products	21
Preliminary Study 2: Archetypal Form	22
Prototype Creation	25
Primary Study	26
Overview	26
Participants	27
Design Evaluation	28
Survey	30
Follow Up Questions	35
Data Reduction	36
RESULTS AND DISCUSSION	40
Overview	40
Preliminary Study 1: Familiar Products	41

Preliminary Study 2: Archetypal Product Form	43
Primary Study: Medium Comparison Analysis.....	46
Primary Study: Full Comparison Method	49
Weighting Values	50
Data Exploration	51
Overall Differences Between Prototype Mediums	53
Relationship Between Object and Prototype Medium.....	53
Relationship Between Attributes and Prototype Medium	56
Selected Qualitative Data	60
Relationship of Valuable Feedback and Prototype Medium	60
Relationship of Prototype and Designer	61
Use of Renderings In The Design Process	61
Population Differences.....	62
Relationship Between Product Selection And Primary Study	
Results	66
Unfamiliar Products.....	Error! Bookmark not defined.
Alternate Archetypal Products	66
Discussion Summary	68
CONCLUSION.....	69
Do Physical Prototypes Or Rendered Prototypes Provide A More	
Effective User Evaluation Of A Product?	69
How Does The Medium Of A Prototype Affect An Evaluator's	
Understanding Of The Prototype?	70
How Does The Medium Of A Prototype Affect User Feedback On	
Certain Design Qualities?	Error! Bookmark not defined.
Impacts On The Design Process	71
Future Work	72
Appendix A	74
References	79

List of Tables

Table 1.47

Table 2.58

Table 3.62

List of Figures

<i>Figure 1.</i> Different levels of prototype realism.	15
<i>Figure 2.</i> Example of product concepts and standardized drawings used for professional designer evaluation.	16
<i>Figure 3.</i> Example question used in Preliminary Study 2.	24
<i>Figure 4.</i> Summary of prototypes.	26
<i>Figure 5.</i> Example comparison of the same product across mediums.	32
<i>Figure 6.</i> Example comparison question of different products across mediums.	34
<i>Figure 7.</i> Example quality check question.	35
<i>Figure 8.</i> How the <i>Full Comparison Method</i> calculated weighted ratings.	38
<i>Figure 9.</i> Graph of average familiarity of 50 common products.	42
<i>Figure 10.</i> Percent agreement among product form selection.	44
<i>Figure 11.</i> Products with the most agreed upon form.	45
<i>Figure 12.</i> The average win and tie percentage of physical and rendered prototypes.	47
<i>Figure 13.</i> Graph of average win and tie percentages by attribute.	49
<i>Figure 14.</i> Average attribute weighting value.	51
<i>Figure 15.</i> Average Rating of Prototypes by Object.	54
<i>Figure 16.</i> The total rating for each prototype medium.	57
<i>Figure 17.</i> Attribute contribution percentage for total ratings.	58
<i>Figure 18.</i> Interaction of Participant Engineering Background and Prototype Medium.	64
<i>Figure 19.</i> Sample Preliminary Familiarity Study 1 Question	74
<i>Figure 20.</i> Archetypal product forms not used in the Primary Study.	75

Effects of Prototype Medium on User Feedback in Product Evaluation

INTRODUCTION

There is an inherent value to a physical thing. By “manipulating them, humans articulate their own relationships with one another” (Ulrich, 2015). Physical things make up our world; we touch them, look at them, and smell them every day. The way these products are designed affects how people interact with them, and influences their feelings towards items they use everyday (Norman, 2004). Think of using a well-balanced, sharp, knife to slice a tomato or clicking the top of a pen. There is something about these and similar products that make them satisfying or pleasant to use. These products and others were intentionally designed to make people feel satisfied when using them. However, products do not simply appear out of thin air, and only having an idea for a product is not enough to develop a functional product

Engineers and designers transform ideas into products that fulfill user needs through the user-centered design. User-centered design requires that all "development proceeds with the user as the center of focus," (Rubin & Chisnell, 2008). A major part of this development process is prototyping and testing. Designers build a representative model or prototype of the unfinished product that approximates some dimension of an unfinished design. Historically, designers have created prototypes out

of anything from foam to wood. However advances in technology it has become easier than ever to create and share designs via Computer Aided Design (CAD) models and realistic computer renderings (Snyder, 2003). Creating physical prototypes with rapid prototyping is also easier than ever, but often the cost is financially prohibitive compared to just printing a picture on paper. In practice, both prototyping approaches are often used in the product design process, but it is unclear if one is more effective than the other at answering unknowns about design concepts.

In the author's own product design experience, one experience particularly stands out. The author showed another design team member a picture of a design for a tool handle grip they were working on and was met with anguish. The fellow team member listed a handful of reasons why the grip could not work and would be uncomfortable. However, the day before, the author showed the same team member a foam carving of the same grip and was met with delight. This pattern repeated several times over the course of the product development process. Intrigued, the author suspected some interplay between the prototyping approach and design evaluation.

The impetus for this research further formed as the author progressed through his mechanical engineering studies. Classes that focused on designing physical parts required pictures of CAD models in reports instead of fabricated physical parts. It quickly became apparent

that a lackluster CAD model and a stellar CAD model would look the same printed on a page. However, when a project required a fabricated part, it was clear which designs successfully translated an idea into a fabricated part and which ones needed more work.

In engineering fabrication classes, students learn why designers often have to make certain design decisions through hands-on experience. For example, tooling and cost constraints explain why fillets should have large radii when possible. The reality of fabrication imperfections guides geometric dimensioning and tolerancing. There is a clear difference between specifying a mating feature on a part drawing, and fabricating that part with a mating feature that fits correctly. All of these concepts can be learned in a classroom, but they are not fully internalized and understood until a student tries to machine a CAD model that cannot physically be machined, or cannot fit two parts together that mated perfectly in CAD.

The learning experience must be meaningful for students to fully understand design principles. In *Educating Engineers For 2020 And Beyond*, Vest (2005) says, "The ultimate goal would be to produce a new breed of engineer who has not only a theoretical understanding of [design] manufacturing, but also practical, hands-on experience."

The same principles apply to designers and the design process out of academia. Can a designer fully understand an idea with only a few

images? Can a user give effective feedback on a concept if he or she does not completely understand it? Does a user need hands-on experience with a product, or does a theoretical understanding suffice? Does using a physical or rendered prototype in the product design process affect user feedback?

Some research has been done on how the medium of a prototype affects the design feedback. This research will examine the potential difference in how valuable a prototype is to a designer depending on the medium of the prototype. Specifically, does a physical prototype or rendered prototype provide a more effective user evaluation of a product, and thus is more useful to a designer? Prototype effectiveness is how well a prototype answers a design question about a concept and will be discussed in more detail below. This research will be useful for engineers and designers by giving them more insight into how evaluators use and interact with prototypes. In academics, students could learn how to use prototypes more effectively, and how to create designs more efficiently. In industry, designers could allocate prototype fabrication resources differently depending on the question they want to answer, or the users they are evaluating. The main questions this research seeks to answer are:

- Do physical prototypes or rendered prototypes provide a more effective user evaluation of a product?
- How does the medium of a prototype affect an evaluator's understanding of the prototype on certain design qualities?

LITERATURE REVIEW

Overview

The design community has studied prototyping in the product design process as a means of gathering user feedback. This review examines the current literature on the purpose of prototypes, effects of prototype fidelity, influences of design methods on prototype evaluation, prototype representation medium, and design evaluation methods. This section will review the interaction between these bodies of work and identify the need for further research.

Purpose of Prototypes

Designers use prototypes to explore physical interactions and usability of their ideas. In the modern product development cycle, designers ideate, test, and prototype their designs to create a refined final product (Eppinger & Ulrich, 2011). Prototypes are built to answer

questions, ranging from “will this work?” to “is this comfortable?” to “would someone buy this?” (Ullman, 2003).

Prototypes are often thought of and referred to by the manufacturing technique used to create them, such as paper, foam, CAD or machining, as shown in Figure 2 (Kiefer, Silverberg & Gonzalez, 2004). This research will use a rapid prototyping method known as 3D printing to create physical prototypes. 3D printing is a process of adding thin layers of a material to create a physical object. Renderings are computer generated realistic simulations of objects, and this research will create renderings using pictures taken from CAD.

Prototypes are also often categorized by fidelity, or the amount they represent the real thing. A very high fidelity prototype may be mistaken for the finished product, while a low fidelity prototype is more akin to a ‘back of the napkin’ drawing. Fidelity is a common term in computer interface design where it is much easier to create a paper prototype for evaluation than a fully coded system (Synder, 2003). Paper prototypes save time and money compared to the massive time and cost associated with creating a full software system. They are often wireframe or full color images of what a piece of software could be. Building physical product prototypes faces a similar trade off in fidelity, time, cost, and effectiveness. Ideally, designers should choose the most cost effective approach, meaning that the prototype can provide valuable information during testing, while still being

built quickly and inexpensively (Dijk, Vergeest & Horváth, 1998). Sketching is another form of prototyping that is valuable in the design process, but this research will not analyze sketch prototypes because “the creation of prototypes requires a set of design skills and time commitment that are a level removed from hand sketching” (Yang, 2005).

Ullman (2003) proposes four categories of prototypes broken down by their function, and the types of question they answer:

- A *proof-of-concept* prototype is used in the initial stages of design to better understand what approach to take in designing a product.
- A *proof-of-product* prototype clarifies a design's physical embodiment and production feasibility.
- A *proof-of-process* prototype shows that the production methods and materials can successfully result in the desired product.
- A *proof-of-production* prototype demonstrates that the complete manufacturing process is effective.

Ullman suggests using a specific type of prototype based on the specific question a researcher is trying to answer. For example, a foam model painted and weighted to look like a finished product could be useful when answering questions about the industrial design, or the ergonomics of a product. A storyboard could answer questions about the steps that users experience when using a product. A functional prototype could look nothing like the final product, but be used to validate new technology or

the essential function of a product. In manufacturing, a production prototype answers the question of how the finished product will look, function, and be made. It is important for engineers and designers to target their prototypes toward answering specific questions, rather than making prototypes because it is simply a prescribed step in the engineering design process (Ullman, 2003).

This research will examine the *proof-of-product* prototype stage, in which a CAD model is created and evaluated for form and production feasibility (Ullman, 2003). A *proof-of-product* prototype could also be called a 'look and feel' prototype that represents the form and appearance of the product; they are often 3D printed, sculpted out of foam core, or rendered 3D models (Ferguson, 1994).

Influence Of Prototype Fidelity On Design Feedback

There is a body of work on how the fidelity of a prototype affects the design feedback, design process, and outcome. Yang (2005) found that prototypes with lower fidelity correlate better with design outcome. They analyzed several design teams' processes and found that the teams that jumped into detailed design too soon had less novel and lower quality final products. In another study, highly realistic computer images did not translate to more understanding of the product as measured by the

number of design features participants could remember in design concepts (Schumann, Strothotte, Laser & Raab, 1996). When Macomber and Yang (2011) examined the use of conceptual sketches for eliciting design feedback, realistic, 'clean' hand drawings were ranked higher than lower fidelity sketches and CAD models, because participants thought the designs were unfinished and their feedback would have more sway on the final design.

Influence of Design Method in the Product Evaluation Process

There are many tools that designers use to gather feedback on designs. This section will examine the use of Computer Aided Design (CAD) to create rendered prototype and physical prototypes in the product design process.

Computer Aided Design (CAD)

The ubiquity of CAD tools in engineering, manufacturing, and product design has made 3D computer models a standard practice to create and convey information among design team members and consumers (Söderman, 2005). Ullman, Wood and Craig (1990) found that the use of CAD early in the design process led to more detailed designs, but fewer overall ideas. In other words, early stage CAD use led to more

depth, but less breadth of ideas. An *in situ* observation of industry practices found that even with advanced CAD systems, designers often took to pen and paper to work out designs, because ideas could be flushed out faster with a pen and paper (Elsen, Darses & Leclercq).

A recent study took a close look at actual industry development projects; Fixson and Marion (2012) sought to discover how ubiquitous CAD use had affected the product development process. To explore this question they performed a longitudinal comparison of two separate, but similar projects, one in 2001 and one in 2009. The 2009 project used significantly higher levels of CAD, and had significant savings in prototyping costs compared to the 2001 project, yet both projects took similar amounts of time and cost similar amounts in the end. The higher use of CAD affected the product development process as well. The 2009 project was transferred to digital CAD models earlier on than the 2001 study, resulting in more design iterations and prototypes in the 2009 study. This was likely because the increased use of CAD made quick iterations easier and more prevalent. However, with the increase in iterations came a decrease in rigid process structure, since all team members could change digital models quickly and easily. The 2009 project did not exhibit any superior performance with labor costs and development time compared to the 2001 project. Fixson and Marion (2012) found that the time and cost increases in the 2009 project were likely caused by the team

jumping directly into detailed CAD design, and solving what were early stage problems in the 2001 project later in the development process. By jumping into digital design too quickly, the 2009 project did not question initial design assumptions and paid the price for it later in the development process by correcting their assumptions at the last minute and negating any potential savings they garnered by using highly iterative design tools. The effortlessness of digital design tools can have unintended consequences.

Renderings

Renderings represent a design with a detailed computer simulation; see Figure 1 (Yang, 2005). They can range from highly detailed environments, such as computer-generated worlds in cinema, to simpler environments, such as a white background. Renderings are relatively easy and cost/time efficient to create and transport compared to a physical part. For this reason they have become almost ubiquitous in all stages of product evaluation (Söderman, 2005). A great deal of research has examined how the fidelity of renderings affects user feedback. Notably, Reid, MacDonald, and Du (2013) found inconsistencies in user opinions when asked to evaluate computer sketches, front/side view silhouettes, simplified renderings, and realistic renderings for product form. However, the variable fidelity of the renderings did not seem to change how much

the reviewer understood the product itself when asked to report design attributes of the represented products.

Physical Prototypes

Engineers and designers use often use physical prototypes to find physical interactions and conflicts with their designs (Ulrich & Eppinger, 2011). Previous research in physical prototyping has looked at the effectiveness of low fidelity prototypes in reducing uncertainty in designers throughout the design process (Gerber, 2009). Houde and Hill (1997) categorized physical prototypes by what the designer can learn from them. He separated them into 'look and feel' prototypes that examine form, color, texture, etc., and implementation prototypes that examine function, strength, materials, etc.

Prototype Representation Medium

Product designers can gain insights from classical art, as the medium of an art piece plays a part in how people view the piece. A painting and a sculpture of the exact same thing can elicit wildly different reactions. Similarly, in the design process, a prototype's medium of representation can affect how people evaluate the design.

Depending on the stage of the design process, a specific design medium can be more useful than any other. In an ideation session a quick 'thinking' sketch can convey enough information for another person to understand the concept and move the project forward (Yang, 2005). A rough, low-fidelity, prototype "can encourage users and other stakeholders to respond in a more creative manner. Especially in the early stages of a project, you might deliberately choose a prototyping method with a rough look to get this benefit," (Snyder, 2007). However, in a late-stage design review, a manufacturing-quality physical model may be necessary to manufacture a product on a large scale. (Ulrich & Eppinger, 2011).

Söderman (2005) compared how sketches, virtual reality representations, and a physical model of a car affected the amount participants understood about the car, see Figure 1. The study indicated that the more realistic representations led to slightly, but not significantly more, participant understanding of the car, meaning there are other factors at play than the degree of product representation realism when consumers evaluate products. Representation medium and fidelity are other factors that could lead to potential differences in evaluator understanding of a product. Like Söderman's (2005) work, this research seeks to explore differences in prototype representation.

Artacho-Ramirez, Diego-Mas, & Alcaide-Marzal (2008) examined the possible limitations of prototype representations and found that as the

representations became more sophisticated, the differences in how participants rated individual prototypes decreased. These studies show that there is not a clear-cut interaction between prototype representation and customer evaluation of the product. Therefore, more research needs to be done to understand the best practices for using design tools and evaluation methods.



Figure 1. Different levels of prototype realism. From left to right the images show a sketch, a 3D rendering, and a full model of a car (Söderman, 2005).

In a very recent study, Häggman, Tsai, Elsen, Honda and Yang, (2015) sought to investigate the relationship between design tool, product qualities, and user evaluation. To answer their questions they asked 18 engineers and product designers in the United States and Belgium to create concepts for a remote control using sketching, foam, or CAD. A team of six design experts evaluated the concepts on qualities such as novelty, form factor, input and interaction. To control for the prototype style differences and presentation medium, a professional industrial designer recreated all concepts as annotated 2D sketches. Figure 2 shows

examples of the concepts created with each design tool, and its respective recreated sketch.

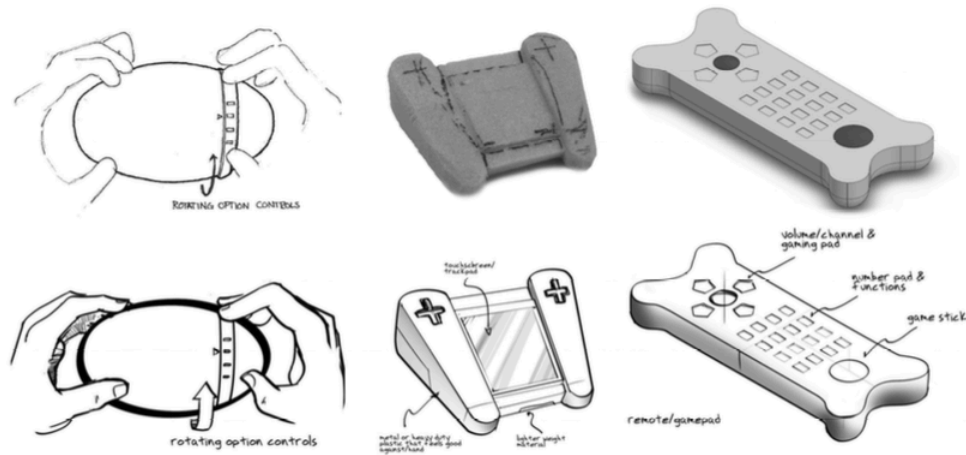


Figure 2. Example of product concepts and standardized drawings used for professional designer evaluation. From left to right: a sketch prototype, a foam prototype and a CAD prototype, and their respective standardized drawings below. (Häggman, Tsai, Elsen, Honda & Yang, 2015).

The researchers used a survey on Amazon Mechanical Turk, a web service where anyone can pay people to complete simple tasks, to evaluate the concepts. They asked 532 people to rate the concepts on their performance, features, and aesthetics, which are three of Garvin's eight attributes of product quality.

The results show that the designers that used foam to create concepts created more overall concepts, more varied concepts, and spent more time creating concepts. Furthermore, looking at concept ratings by

design attribute, foam prototypes account for 57% of the top 20 most comfortable concepts, 62% of the 20 most creative concepts, and 56% of the top 20 most aesthetically appealing concepts. Concepts created using CAD were rated similarly to current offerings of remotes. Häggman et. Al (2015) found that the use of CAD early on in the concept generation phase of the development process resulted in three fold more time spent creating each concept, but half as many completed concepts, compared with either sketching or foam prototyping. In the study, the medium of the design tool led to categorical differences in the designs. This shows that there is some interaction between the design tool and the outcome. Additionally, the body of research on the product design process points to interactions between prototype medium and design outcome that may not be readily apparent, yet hold consequences for the process and outcome.

Design Evaluation

There are many methods used in product evaluation. Wickens, Gordon and Liu (2004) outline several methods for Human Factors product design evaluation and best practices for using prototypes in the evaluation process. They write that anyone can evaluate prototypes, but an experienced evaluation team will target prototypes to specific types of users to gain the most insight into the design concept. Most commonly, a

moderator sits down with a participant and goes through a series of questions and tasks with the prototype (Krug, 2010). These usability evaluation sessions are often recorded for members of the design team to watch and learn from the session. Sometimes prototypes are created and evaluated so quickly that only the design team sees them. Informal testing of quick prototypes is very common especially in early stages of the design process (Kudrowitz & Wallace, 2013).

Research Gap

After examining the current literature on the purpose of prototypes, effects of prototype fidelity, influences of design methods on prototype evaluation, prototype representation medium, and design evaluation methods, -- as well as how they interact with one another -- we seek to propose new research that can fill gaps in the current body of work on product design and the use of prototypes in the design process.

Past research has shown that differences in prototype presentation, both the medium and fidelity, can affect users' feedback when evaluating concepts. However, with emerging technologies such as low cost 3D printing, it is easier than ever to create physical prototypes. At the same time, it is easier than ever to create realistic 3D renderings with the

immense computing power in smartphones and low cost computers. The design process requires efficient use of time and money, while allowing the freedom to explore and evaluate new ideas.

Specifically, Fixson and Marion (2012), Häggman et al (2015), and Söderman's (2005) work highlights the effects of prototype medium and the use of digital design tools in the product development process. We will build on their work using ideas and methods based on past research, while employing modern research and design tools.

Much of the work on physical product prototype evaluation has been focused on variations in prototype fidelity, but not the medium. Looking beyond classical design evaluation methods, we find there are many unanswered questions about how the prototype medium affects how much a user evaluating a prototype understands about the product it represents. This research is particularly relevant because the majority of research has focused on the best processes for designers to create more ideas of higher quality, while comparatively less research has been done on user feedback on prototypes. We seek to find the potential interaction between prototype medium and user evaluations.

METHODS

Overview

The research team conducted two large sample surveys to choose which products to prototype. Once four products and forms were chosen, the research team created physical and rendered prototypes of each product.

In the primary study, sixteen participants evaluated eight prototypes in a user test setting. Then the participants completed a questionnaire asking them to compare each prototype against every other prototype. Finally, the research team asked each participant qualitative questions about their experience and thoughts about each prototype.

Preliminary Studies: Product Selection

The first step of designing the study was to identify the products that participants would evaluate. The search for familiar products as part of the larger experiment will be referred to as the 'preliminary study.' This study is not the main focus of this research, but is necessary to explain how the research team chose which products to use in the primary study.

The research team decided to use familiar products for design evaluation in the primary study because the layperson has the ability to make design judgments on products that they interact with on a regular basis. Furthermore, using familiar, everyday objects could generalize this

research to the broader field of consumer product design. Asking a participant to give feedback on an obscure coffeemaker, for example, could skew the results, or provide no results at all because it is unlikely a randomly selected participant could effectively evaluate an object that they have no experience with. Previous research has looked at multiple design interpretations of a single object. Given the goals of this research, the research team felt it was necessary to have concrete reasons why they chose the products used in the study .

Preliminary Study 1: Familiar Products

To find familiar products, the research team created a survey that asked participants to rate their familiarity with common, everyday products. The research team chose 50 items from the best-selling products on Amazon.com as of May 2015 for the survey. Some products included in the preliminary study were a knife, a frying pan, and a doorknob, among others. The survey, which was created with Google Forms, showed participants a picture of the product with all branding removed and asked them to rate their familiarity with the object on a scale from one (no familiarity with the product) to seven (very high familiarity with the object). To remove any ordering or exposure bias, the questions were randomized so that each participant rated the products in a different order. To ensure validity, the survey asked the participants to give a name to the product. If

a participant rated an object as very familiar, but did not know the name of the object, then that rating was not considered valid.

The survey was distributed through Amazon Mechanical Turk to gather a large number of responses. Amazon Mechanical Turk is a web service that allows anyone to anonymously complete tasks for pay, and is often used in the social sciences to easily gather a large and varied sample of responses (Buhrmester, Kwang, & Gosling, 2011). Participants were limited to Mechanical Turk users that were Americans with over a 97% task approval rate. Approval rate is how often a worker's work is considered acceptable. 100 respondents completed the survey, and after responses were checked for validity, all 100 responses were accepted. The products with the highest average familiarity – with a score of over six on the one to seven scale – were considered highly familiar products. Of the 50 products in the survey 40 had an average score above six. The top 18 highest rated objects were selected for further analysis. These products had an average familiarity ranging from 6.7 to 6.95 out of 7. Complete results from study can be found in the Results and Discussion Section

Preliminary Study 2: Archetypal Form


After the familiar products were chosen, the research team had to identify the form of each product, which describes how a product looks


and feels. In competitive markets, product form can differentiate successful products from unsuccessful ones. To select the specific form, the research team created a survey in Qualtrics, an online questionnaire creation tool, that asked participants to choose a product form that best matched their idea of said product.


Participants were asked to picture one of the 20 highly familiar products in their mind. Then they were shown images of four common versions of the target product and were asked to select which image best represented the product form they had pictured. If none of the products matched their idea of the target product they could select, “None of the above products match my idea of the target product,” or if they thought there was no difference in the product forms they could select, “I don't think there is a difference in the above products.”


For example, a participant was asked to picture a permanent marker in their mind. On the next page they were shown four images of different forms of a permanent marker, as shown in Figure 3. They selected the form that best matched their idea of a permanent marker.

Which of the following best represents your idea of a **Permanent Marker**?









☐ None of the above products match my idea of the target product

☐ I don't think there is a difference in the above products

Figure 3. Example question used in Preliminary Study 2. These questions were used to select archetypal product forms. Participants selected which form best match their idea of the target product.

The second preliminary study was also distributed through Amazon Mechanical Turk. The same participation requirements were used in this study as the first preliminary study: participants were limited to Mechanical

Turk users in the United States with approval rates above 97%. In total, 100 responses were collected.

The research team analyzed the data to find the product forms that best matched the target products. For each target product, the response that was selected the most was considered that product's archetypal representation, see Results and Discussion.

However, the research team had additional selection criteria for choosing which archetypal products would be used in the primary study. First, the archetypal product form had to have been chosen by over fifty percent of respondents, and second, it had to be feasible to prototype. Using these selection criteria, the research team chose a doorknob, a permanent marker, a drinking glass, and a lightbulb for the primary study.

Prototype Creation

For each of the products chosen, CAD models were created using Onshape, an online CAD modeling software. The CAD models were exact copies of the archetypal product forms.

To create the physical prototypes that would be used in the primary study, the research team 3D printed each CAD model with an Objet 500 Connex 2 using the Veroclear material with a matte finish. The 3D printed objects were then painted with a matte grey spray paint.

Renderings were created by taking screenshots of the Onshape workspace. The CAD models were given the same grey color as the physical prototypes. The object renderings were presented at a one to one life size scale at an isometric viewing angle. The rendered prototypes were printed in single-sided color on letter size paper with one prototype per page. See Figure 4 for pictures of all prototypes.

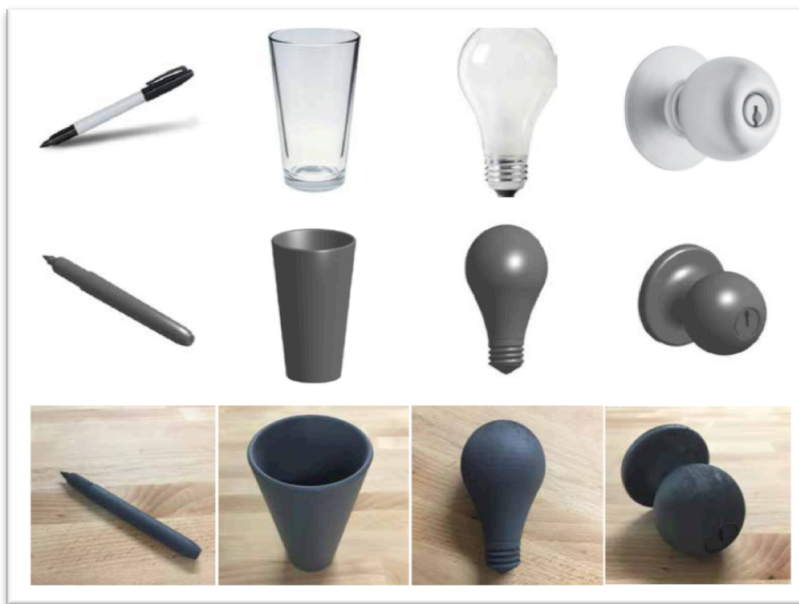


Figure 4. Summary of prototypes. The first row shows the chosen form, the second row shows the rendered prototypes, and the third row shows the physical prototypes.

Primary Study

Overview

The main questions this research seeks to answer are:

- Do physical prototypes or rendered prototypes provide a more effective user evaluation of a product?
- How does the medium of a prototype affect an evaluator's understanding of the prototype on certain design qualities?

This research seeks to answer these questions with an experiment in which participants evaluate prototypes in a user evaluation study, compare each prototype based on certain design attributes, and answer questions about their thought processes.

Participants

Participants for the primary study were recruited through Tufts University Human Factors classes and on campus postings. Participants recruited through Human Factors classes received class participation credit for participating in the study. The research team did not have any conflicts of interest with participants, classes, or professors that awarded credit. Participants recruited through campus postings did not receive any compensation.

The 16 people that participated in the study ranged in age from 18 to 25 years old. Of the 16 people, 15 were undergraduate students while one was a post-baccalaureate student. Participants were given the option

to report their gender, ethnicity and academic major, and all participants did so. Nine were men and seven were women. 13 participants were white, two were Asian, and one was Hispanic. Six were human factors majors, two were mechanical engineering majors, and the other eight were an assortment of liberal arts majors.

Design Evaluation

As previously discussed, design evaluation is an important part of the engineering design process, and user design evaluation was a major part of this research.

The design evaluation took place at the Tufts University usability lab. The primary investigator observed the experiment from behind a one-way mirror in the observation room. To remove ordering bias, the research team randomly assigned participants to evaluate the physical or rendered prototypes first. The order that the individual prototypes were presented was also randomly selected prior to each study using a random number generator.

The moderator, an undergraduate in the Human Factors Engineering program, greeted each participant and led him or her into the usability lab. The only thing present in the room was an informed consent document. All prototypes were kept out of sight until the participant

evaluated them. The moderator walked the participant through the study and the informed consent form. Once the participant agreed to the informed consent document, the experiment began.

First the participant answered yes/no priming questions about the product the prototype represented (see Appendix A for the full list of questions). Then, the moderator presented the participant with a prototype and gave him or her 20 seconds to internalize and interact with it before asking questions. The moderator asked the participant:

- Can you give me your first impression of this product?
- What's a typical use of this product?
- What do you like about this product?
- What do you not like about this product?
- Can you give me your overall impression of this product?
- Do you have anything else to add?

Once the participant finished giving design feedback, the moderator took the prototype away and put it out of sight. The moderator then presented the participant with two yes/no priming questions for the next product of the same medium (either rendered or physical). This process repeated until the participant had evaluated all the prototypes in that medium. Then the moderator presented the participant with the other medium prototypes in the same order with more yes/no priming questions.

The primary investigator observed and recorded the participants' design feedback for any useful or revealing comments.

After the participant had evaluated all of the prototypes in both mediums, the moderator gave the participant an iPad with a preloaded Qualtrics survey, detailed below, and left the room. Once the participant completed the survey the moderator and the primary investigator came back into the observation room and asked the participant qualitative follow-up questions, which will be detailed below, about his or her thought process in the study.

Once the research team finished discussing the survey with the participant, the participant was given their compensation and allowed to leave the room. Each participant took approximately 45 minutes to complete the whole process.

Survey

The survey was a major data collection tool in this experiment. The survey asked participants to compare the design attributes of every prototype against every other prototype. The research team chose from Garvin's Eight Qualities of Product Quality to describe different attributes of the prototypes. The eight attributes are perceived quality, features, performance, aesthetics, reliability, serviceability, durability and conformance to standards (Garvin, 1984). An important consideration in

choosing these criteria was whether it was possible for participants to make judgments on attributes based on early stage prototypes. It was decided that it would be difficult to judge reliability, serviceability, durability and conformance with early stage prototypes, so only the perceived quality, features, performance, and aesthetics were analyzed. This line of thinking echoes how Haggman et al (2015) evaluated early stage design concepts.

Since different people prioritize certain product attributes above others, the survey began by asking participants to weight how much they care about each of the four attributes relative to the others, with the total weight of all four equaling 100. If a participant cared about all the attributes equally, he or she could give each attribute a weighting of 25. Similarly, if he or she only cared about aesthetics, he or she could give aesthetics a weighting of 100 and the other three attributes a weighting of zero.



After completing this step, the survey presented pictures of two prototypes and asked participants to compare the two prototypes on the perceived quality, features, performance, and aesthetics with the following questions:

- Which prototype would you be more likely to buy (assuming they were similarly priced)?
- Which prototype has more uses?

- Which prototype would be more comfortable to use?
- Which prototype is more aesthetical appealing (looks, feels, tastes, smells, or sounds better)?

These questions are similar to design evaluation questions used in Haggman et al (2015). For each question, the participant could say that the two prototypes were equal, or that one was better than the other. The question order was randomized so it would not be the same for each comparison. Figure 5 depicts a sample question from they survey.

Select which prototype best answers the following questions:

Light Bulb Physical Prototype
Light Bulb Rendered Prototype

Select which model best answers the following questions:	Physical Prototype	Somewhat more	Equal	Somewhat more	Rendered Prototype
Which prototype would you would be more likely to buy (assuming they were similarly priced)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Which prototype is more aesthetically appealing (looks, feels, tastes, smells, or sounds better)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Which prototype has more uses	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Which prototype would be more comfortable to use	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



Figure 5. Example comparison of the same product across mediums.

The research team posed the prototype comparisons in a specific order. The first four questions were always comparisons of the same product across mediums in a random order (e.g., Permanent Marker Physical vs. Permanent Marker Render, Glass Physical vs. Glass Render, etc.). This decision was intentional, since these comparisons specifically

looked at different prototype mediums of the same product, which is the main focus of this research. The research team thought that participants could lose focus later in the survey and wanted to ensure that these comparisons were accurate by setting a baseline for comparison. The research team did not want participants confused at first by comparing two rendered prototype to each other.

After the four initial comparisons, the survey moved into all the other possible combinations of prototype comparisons presented in random order (e.g., Light Bulb Render vs. Permanent Marker Physical, Door Knob Physical vs. Glass Physical). Figure 6 shows an example question. The research team felt it was valuable to compare each prototype to every other prototype because it can measure for relative differences between prototypes. By comparing every possible combination of prototypes the research team could learn how prototype medium interacts with different products and attributes. Each prototype was compared seven times in the survey -- once to every other prototype -- for a total of 28 comparisons.

Select which prototype best answers the following questions:



Light Bulb Physical Prototype Door Knob Rendered Prototype

	Select which model best answers the following questions:				
	Physical Prototype	Somewhat more	Equal	Somewhat more	Rendered Prototype
Which prototype would you would be more likely to buy (assuming they were similarly priced)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Which prototype has more uses	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Which prototype would be more comfortable to use	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Which prototype is more aesthetically appealing (looks, feels, tastes, smells, or sounds better)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 6. Example comparison question of different products across mediums.

The research team built in two quality check questions in the survey to make sure participants were paying attention and providing quality data, see Figure 7. The two quality check questions prompted the participant to select a specific answer for that question instead of showing them two prototypes.

People in the Tufts University Human Factors program participated in pilot studies to ensure the survey measured prototype effectiveness. These pilot studies asked five people familiar with design evaluation to evaluate the prototypes and complete the survey. Then they were asked if they thought the survey actually analyzed prototype effectiveness. Three

of these people were experienced human factors engineering professionals. It was decided that the survey effectively measured prototype effectiveness, and that participants could effectively make comparisons between prototypes.

Select "Physical Prototype" for all questions:

	Select which model best answers the following questions:				
	Physical Prototype	Somewhat more	Equal	Somewhat more	Rendered Prototype
Which model would you be more likely to buy (assuming they were similarly priced)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Which model would be more comfortable to use	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Which model is more aesthetically appealing (looks, feels, tastes, smells, or sounds better)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Which model has more uses	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 7. Example quality check question.

Follow Up Questions

In user evaluation, it can be hard to fully understand what users are thinking, so evaluators often ask many follow up questions to fully ascertain users' opinions (Krug, 2010). These questions helped the research team understand more about the participants' thought process, and tease out differences in user preferences for physical and rendered prototypes. The research team asked each participant the following questions:

- Which prototype did you give the best feedback on overall?
- Which prototype did you like the best?

- If you were a designer, which prototype do you think would give you the most valuable information?
- In general, did you like the physical or rendered prototypes more?
- In general, did you give more valuable feedback with the physical or rendered prototypes?
- If you were an engineering professor and assigned a project to design a consumer product, which prototype would you give the highest marks to?

Data Reduction

The raw survey data had to be transformed into a useful form to show relationships and differences between the different prototypes and mediums. The research team reduced the data in two different ways. The first, called *Medium Comparison Method*, assumed that participants were only able to compare different mediums of the same product. The second, called the *Full Comparison Method*, assumed that all participants fully understood each question and were able to compare every prototype to every other prototype.

MEDIUM COMPARISON METHOD

In user testing, it is always possible that users do not answer questions the way researchers intended, so it is valuable to collect data in ways that check for internal validity (Krug, 2010). From the beginning, the research team designed the survey so it could provide useful data comparing prototype medium of the same product. As discussed above, the first four comparisons in the survey were of the same product of different mediums (e.g. Permanent Marker Physical vs. Permanent Marker Render, Glass Physical vs. Glass Render, etc). In the *Medium Comparison Method* the research team assumed that the survey did not work as planned, and that participants could only effectively judge prototypes of the same product on a simple scale. It is important to note that the research team balanced the number of products that would be compared with the total number of comparisons a participant would have to make. With four products participants would have to make 28 comparisons, and with five products they would have to make 45 comparisons. Since the research team was trying to not make the survey incredibly long they thought four products gave the best balance of direct prototype comparison and total number of comparisons.

The data reduction process for the *Medium Comparison Method* was much simpler than the *Full Comparison Method*. The research team only used the direct comparisons of the same product, and did not take user attribute weightings into account at all. In each comparison, a

prototype could only win, lose, or tie for each of the four questions. A prototype won a comparison if a participant had any preference for it, and the prototypes tied if the participant chose 'Equal.'

The win/lose/tie percentage was calculated based on how often a prototype won/lost/tied in relation to the total number of times it was compared. The four prototype pairs were compared 16 times in total.

FULL COMPARISON METHOD

For the *Full Comparison Method*, the research team assumed that every participant completely understood the survey and that all parts of the survey correctly measured what was intended. This section and Figure 8 detail how the research team found weighted ratings of each prototype.

	Comparison Questions					Numerical Value	Weight Value	Weighted Question Value		
	Physical Glass	Equal	Equal	Equal	Rendered Glass			Physical Glass	Rendered Glass	
Performance	●	○	○	○	○	→ 2	20	40	0	Performance
Aesthetic	○	○	●	○	○	→ 0	10	0	0	Aesthetic
Quality	○	○	○	●	○	→ 1	30	0	30	Quality
Features	○	○	○	○	●	→ 2	40	0	80	Features
								<u>40</u> ⁺	<u>110</u> ⁺	Complete Rating Value

Figure 8. How the *Full Comparison Method* calculated weighted ratings. This figure shows an example of how weighted ratings were calculated for each prototype.

First, the research team assigned values to each prototype rating for how far each rating was from equal. If a participant selected the Glass Physical Prototype as the better prototype for a question, then the research team considered that the Glass Physical Prototype ‘won’ that comparison. If a participant selected ‘Equal’ then neither prototype ‘won’ that comparison.

The research team assigned numerical values to the ratings. An equal rating gave both prototypes a zero. The winning prototype was given a two, and the losing prototype was given a zero. If ‘Somewhat More’ was selected, then the research team still considered the respective prototype to have won and assigned it a value of one and the losing prototype a value of zero.

Using the weighting values that respondents assigned to attributes at the beginning of the survey, the research team calculated a weighted ranking score for each prototype called the *Complete Rating Value*, which measured how effective a prototype is. The research team calculated the *Complete Rating Value* by multiplying the weighting value and the numerical value for each comparison to create a weighted question value. Then the research team summed the weighted question values to find the *Complete Rating Value*. This same procedure was repeated for every comparison. See Figure 8 shows how the research team performed these calculations. It is important to note that this measure is only relevant for

comparing prototypes in this research, and would not be applicable to other prototype evaluation experiments.

RESULTS AND DISCUSSION

Overview

After completing the user evaluation experiment in the Primary Study, the research team analyzed the quantitative and qualitative data to answer the questions posed in this research. This section will discuss the results of the Preliminary Studies, the Primary Study and the main findings of this research.

Both data from the *Medium Comparison Method* and the *Full Comparison Method* were analyzed in hopes of answering the main research questions. Differences between objects, attributes and mediums were explored. Selected responses from the qualitative question portion of the experiment are reported to understand how and why participants made their choices.

The main questions this research seeks to answer are:

- Do physical prototypes or rendered prototypes provide a more effective user evaluation of a product?
- How does the medium of a prototype affect an evaluator's understanding of the prototype on certain design qualities?

Preliminary Study 1: Familiar Products

A significant portion of this research included the selection of products that participants evaluated. The research team wanted to ensure that participants could effectively provide feedback on the prototypes by choosing familiar products. Data from 100 Mechanical Turk workers showed which products people were most familiar with. Validation questions were checked to ensure participants were actually familiar with the products they rated as highly familiar. For any individual rating of 5 or above, the name the participant gave the product was checked for accuracy. Figure 9 shows how familiar people were with common products. The research team decided to use the top 18 most familiar products in the next phase of the study. These 18 products had average familiarity scores above 6.7. These results were not surprising because the products are commonly found in most Americans' day-to-day lives.

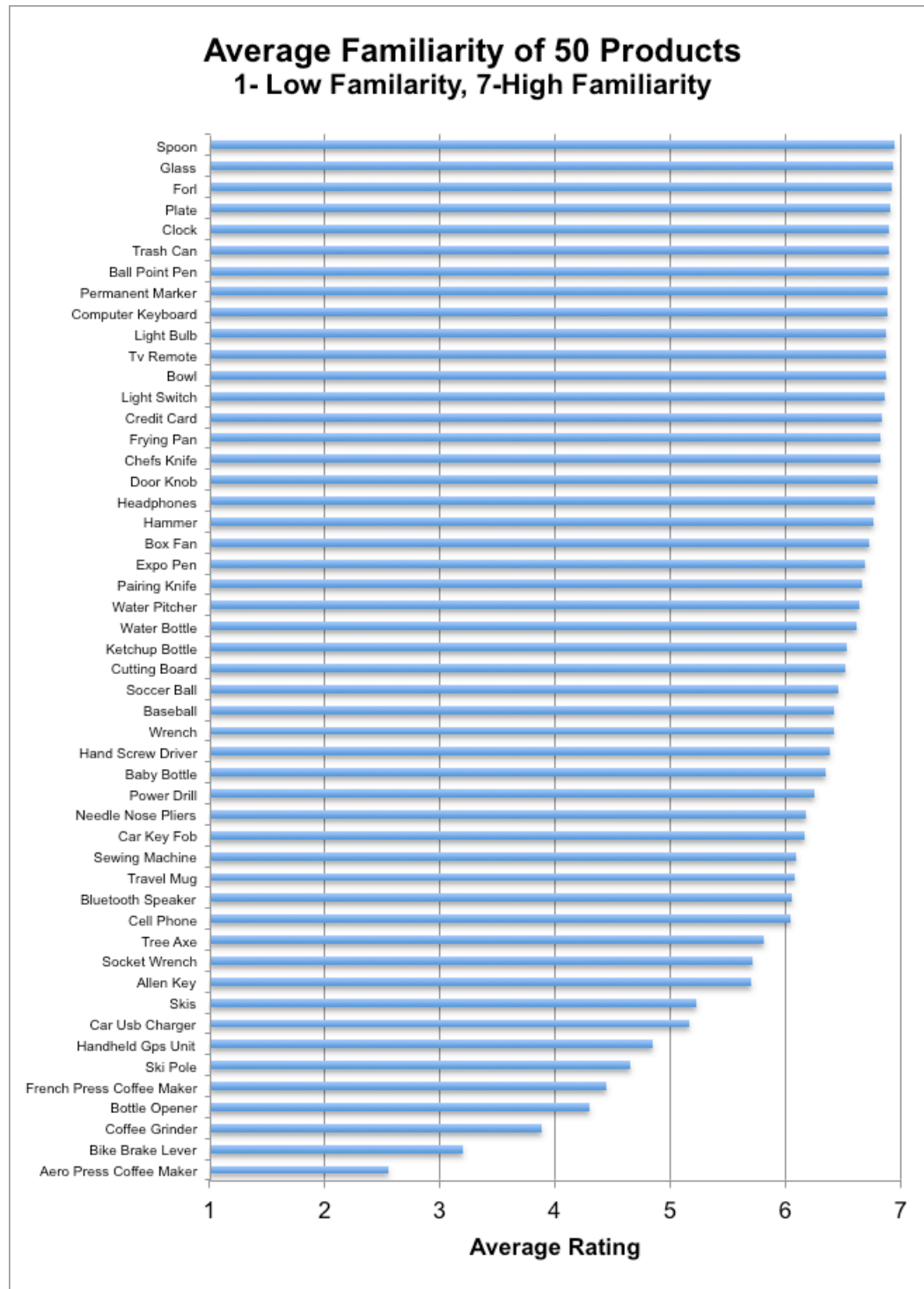


Figure 9. Graph of average familiarity of 50 common products. People rated their familiarity with these common products, 1 is low familiarity, 7 is high familiarity.

Preliminary Study 2: Archetypal Product Form

Since there are many different types of common products, it was necessary to survey people in order to decide on the specific product form. The survey asked 100 Americans to picture the 18 familiar products from the first preliminary study in their mind and choose which archetypal form best matched their mental model of that product. The results show that some product forms were considered more representative than others. Figure 10 shows how much people agreed on the most archetypal products forms.

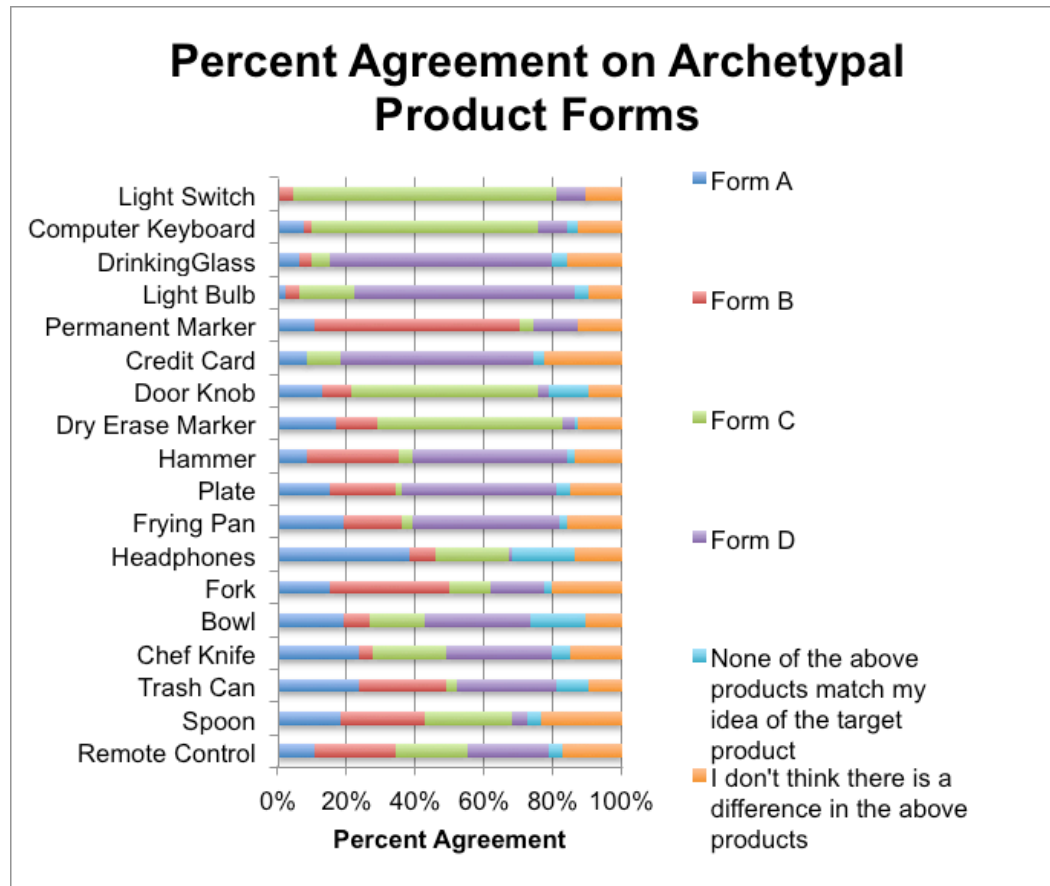


Figure 10. Percent agreement among product form selection. This figure shows the percent that each choice was selected as the most archetypal product form.

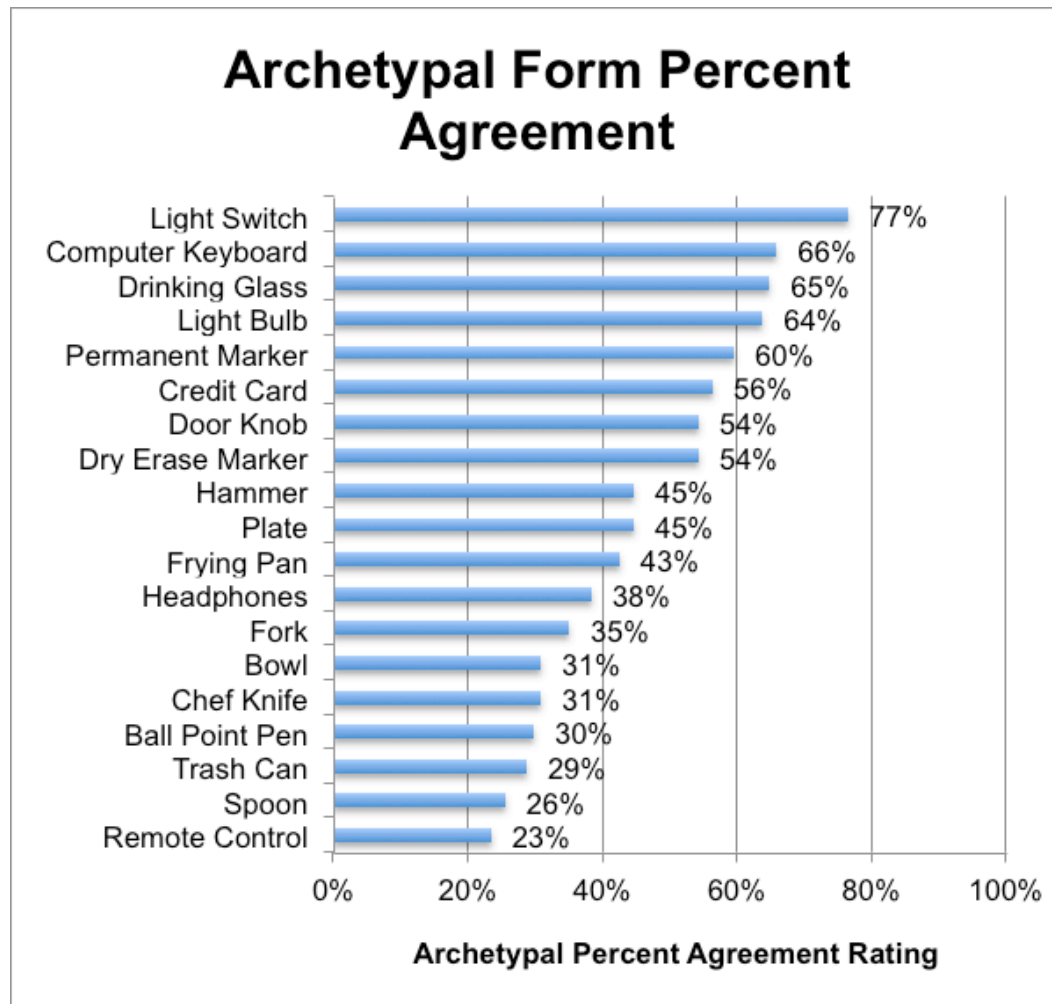


Figure 11. Products with the most agreed upon form. Specific forms of these products were the most agreed upon forms.

The research team used these results to pick specific product forms for the primary study. In order meet the selection criteria, more than 50% of respondents had to agree that a specific form was the archetypal representation of that product, see Figure 11. Eight product forms met the selection criteria. The eight products were a Light Switch, Computer Keyboard, Drinking Glass, Light Bulb, Permanent Marker, Credit Card, Doorknob, and Dry Erase Marker.

Selecting the products that would be used in the primary study was a carefully made decision. Immediately, the credit card was ruled out because credit cards are rarely purchased, and are basically just one object with different branding. (The reason it was included in the surveys to begin with was because prepaid debit cards were a common product on Amazon.com). Only one of either the Permanent Marker or the Dry Erase Marker could be included because the objects were very similar, and it was redundant to include both. The Permanent Marker was chosen because it was considered a more representative form than the Dry Erase Marker. The Light Switch and Computer Keyboard were considered more difficult to prototype than the other remaining objects, and were omitted. The products selected for the primary study were the Doorknob, The Permanent Marker, Light Bulb, and Drinking Glass. These products require frequent human interaction through holding, turning, gripping, or writing. It is likely that participants could evaluate these product forms because participants likely have some idea of or experience with these specific products. Implications of using these specific products in the research are discussed in detail below.

Primary Study: Medium Comparison Analysis

Analyzing only the *Between Medium Comparison* data shows how participants responded to physical and rendered prototypes, and whether they thought there was a difference in how useful the prototypes were. Figure 12 shows that physical prototypes were rated higher more than twice as much as rendered prototypes.

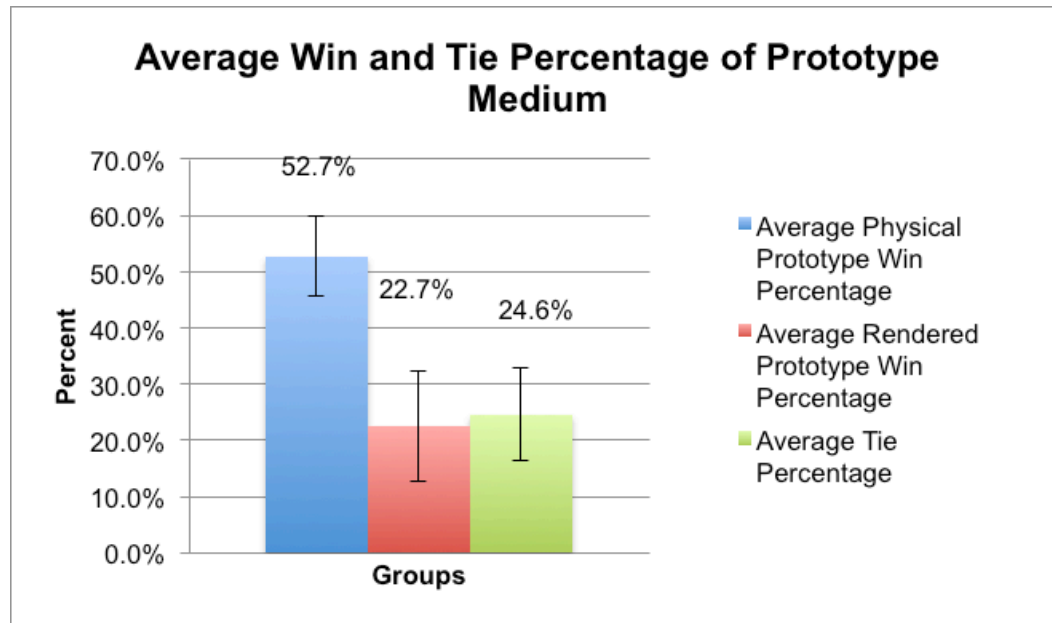


Figure 12. The average win and tie percentage of physical and rendered prototypes. Error bars show +/- 1 standard error. Means are reported for each group.

An independent samples t-test found there was a statistically significant difference between the win percentages of physical ($M=52.7$) and rendered ($M=22.7$) prototypes, $t(16)=5.04$, $p<.001$, however the effect size was small $R^2=.018$.

The four design attributes affected prototype ratings as well. There was a significant difference in the win percentage of physical and rendered

prototypes for performance and feature ratings. Figure 13 and Table 1 summarize these findings. Interestingly, the smallest difference was found when judging aesthetics. One explanation could be that renderings created a more idealized and better looking prototype so more participants rated them higher for aesthetics.

Table 1					
<i>Summary of attribute differences between mediums.</i>					
<i>*Note significance was $p < .0125$ instead of $p < .05$ because the data was split into four groups.</i>					
Attribute	Physical Prototype Average Win Percentage	Rendered Prototype Average Win Percentage	Test Statistic	Significance	Significant Difference ($p < .0125$)
Aesthetics	53.1	40.6	$t(16)=1.01$	$p=.350$	No
Features	48.4	3.1	$t(16)=6.66$	$p<.001$	Yes
Performance	60.9	18.8	$t(16)=4.24$	$p=.002$	Yes
Quality	48.4	28.1	$t(16)=1.83$	$p=.118$	No

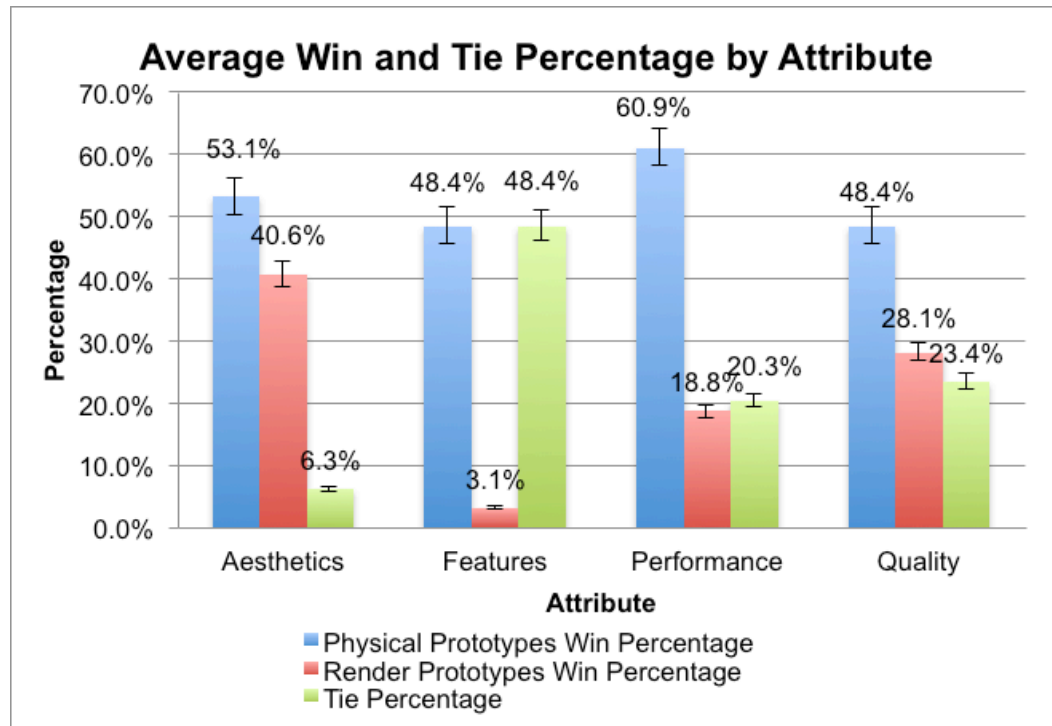


Figure 13. Graph of average win and tie percentages by attribute. Error bars are ± 1 standard error. Means are reported for each group.

These results show that, in general, physical prototypes better represent products, and that physical prototypes are best at answering design questions related to performance. Rendered prototypes were almost universally rated worse for features. One explanation could be that participants found more uses with the physical prototypes when they interacted with the model, resulting in a more complete experience with the prototype.

Primary Study: Full Comparison Method

Data from the full comparison was analyzed to find the potential effects of prototype medium and product on design feedback. The goal was to find whether specific prototypes and attributes elicited more complete feedback from users. The full comparison data provided a more granular look at the differences between mediums because it showed not just which prototype was better, but by how much.

Weighting Values

The reported design attribute weighting values were explored. On average, performance was the most important attribute participants considered when evaluating prototypes as seen in Figure 14. Features were the least important, likely because Garvin describes Features as non-essential to the core use of the product.

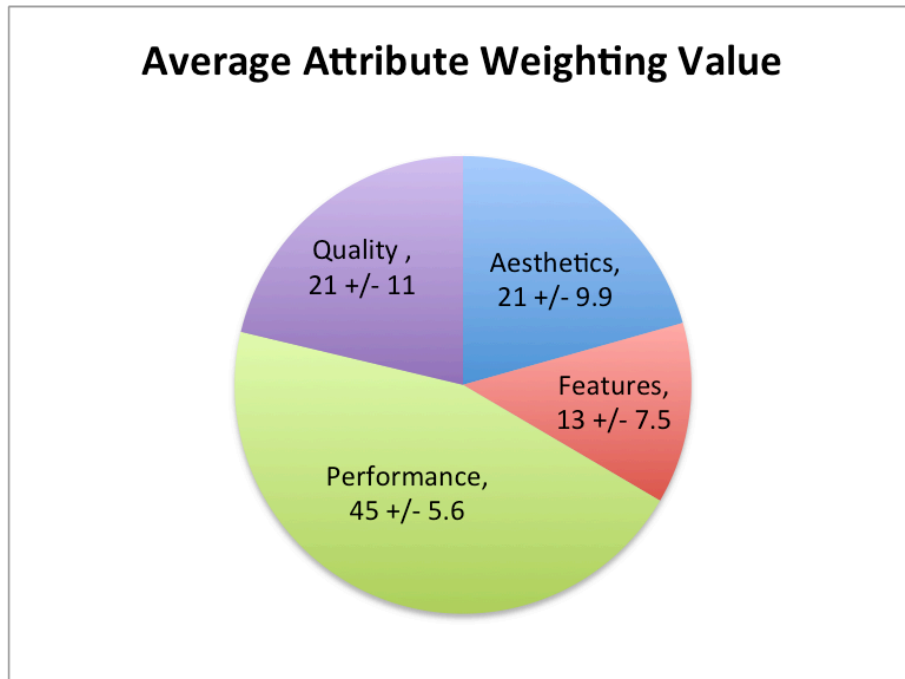


Figure 14. Average attribute weighting value. Standard deviation is shown.

Data Exploration

Once the raw data had been manipulated into a useful form, some preliminary analysis was performed. The analysis showed that the full comparison data was highly non-normal. Upon further inspection, the data was found to follow a natural lognormal curve. The skewness and kurtosis values describe how normal a data set is. Smaller values indicate a more normal distribution. The skewness and kurtosis values went from highly significant to non-significant with the natural log transformation (20.6 to 1.81 and 9.08 to 1.27, respectively). The data is natural log normally

distributed likely because the survey data transformation showed how much each prototype was considered better than equal to its counter part.

After additional examination of the *Full Comparison* data and follow-up questions, it was deemed that there could be significant systematic flaws in the survey. Namely, participants could not effectively compare every prototype to every other prototype. Four participants said that they were unsure how to compare two rendered prototypes of different objects and that they “sort of guessed.” However, participants said they did not have any problems comparing prototypes of the same product in different mediums. The survey was designed to allow for this contingency because the first questions participants answered were the direct comparisons.

After deciding that the *Between Medium* product comparisons were the most consistently valid measures of prototype medium effectiveness, the research team chose to only use the four direct prototype comparisons in further analysis because the chance of valid data was higher. The scaled ratings computed for the *Full Comparison Method* were still used in all analysis so the relative differences between prototypes could still be analyzed. This subset of the *Full Comparison* data was referred to as the *Four Full Comparison* data. The *Four Full Comparison* data was found to follow a natural lognormal distribution. The skewness and kurtosis was found to be non significant in the natural log normal transformed data, $z=.411$ and $z=1.448$, respectively. The natural lognormal data fulfilled the

assumptions needed to use normalized statistical tests, such as t-tests (Rosner, 2011). Further analysis was performed with confidence that the data effectively measured prototype usefulness.

Overall Differences Between Prototype Mediums

Analysis was performed with the more conservative data survey that was thought to be more valid in measuring prototype usefulness. An independent samples t-test showed that there was a statistically significant difference of the average rating of usefulness between rendered prototypes ($M=32.6 \pm 20.9$) and physical prototypes ($M=50.5 \pm 32.1$), $t(16)=3.874$, $p<.001$. However, the effect size was relatively small $R^2=.10$.

The research team is confident that physical prototypes are more useful in user testing. Differences in attributes, objects, and participants were further explored to find why a difference between prototype mediums exists.

Relationship Between Object and Prototype Medium

The ratings for individual objects across mediums showed that all physical prototypes were considered more useful for each product. Figure 15 shows the interaction of objects and ratings across mediums. The smallest difference between mediums was the Drinking Glass. Insight in this difference was found in the follow-up questions.

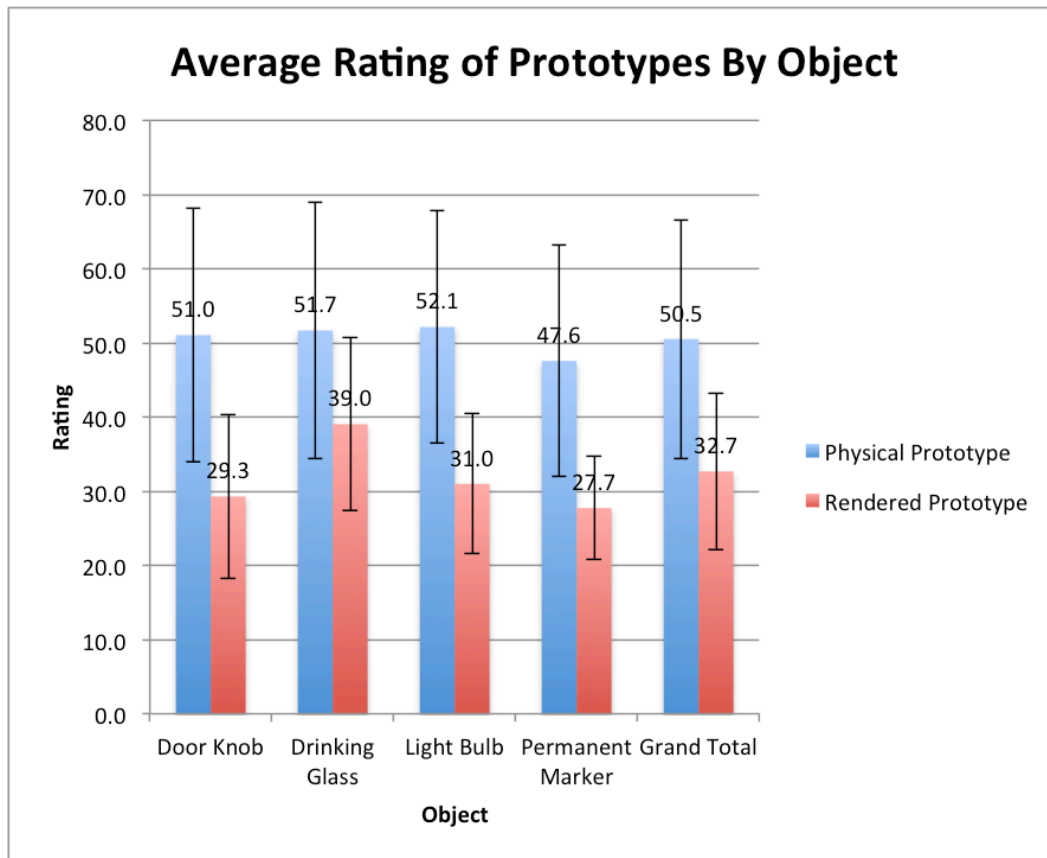


Figure 15. Average Rating of Prototypes by Object. Bars show +/- Standard error. Numbers are average rating.

When asked about their decision making process for the Drinking Glass, participants said they didn't like the texture inside the glass, calling it "dirty," "unclean," and "unhealthy to drink out of." There were remnants of the 3D printing process inside the bottom of the glass that made the surface look textured and dirty, which is generally not how people want drinking glasses to look. One person even said, "I know this is how 3D printed things look, but it just looks gross on the inside, and I can't get past that." On the other hand the renderings were perfect, idealized versions of

the product, so the surface finish and fabrication remnants could not affect user evaluation.

The largest differences between prototype mediums were found in the Doorknob (21.7) and the Light Bulb (21.1). Both of these products require standardized fitments to work, and the Doorknob requires specialized user interaction, in that it needs to be grasped and twisted to operate. People could have thought physical prototypes were more useful for products that need interaction with other fabricated parts. Differences between mediums in the Permanent Marker (14.9) and the Drinking Glass (12.7) were smaller than the other products. These products are often used independently of other objects.

The survey's original intent was to compare all prototypes against every other prototype to find fine differences across medium, attribute and object, but no significant conclusions can be drawn, because the research team lacked complete confidence in the *Full Comparison*.

These findings follow the idea that prototypes should answer specific questions. Neither an early stage 3D printed, nor a rendered prototype can answer how hygienic a product would be, or how the production material feels. Early stage user opinions can be tainted by confounding factors, such as remnants of the fabrication process, that were not intended for evaluation. For this reason, it is important for user

evaluation practitioners to target prototypes to specific questions and help users overcome misgivings not related to the main evaluation purpose.

Relationship Between Attributes and Prototype Medium

Analysis for each attribute was performed to understand if one prototype medium was more effective than another at eliciting feedback for a specific product design attribute. The sum of the *Computed Rating Values* of each medium is shown in Figure 15. Across the board, the physical prototypes had higher ratings, but the ratings' make-up showed some interesting differences.

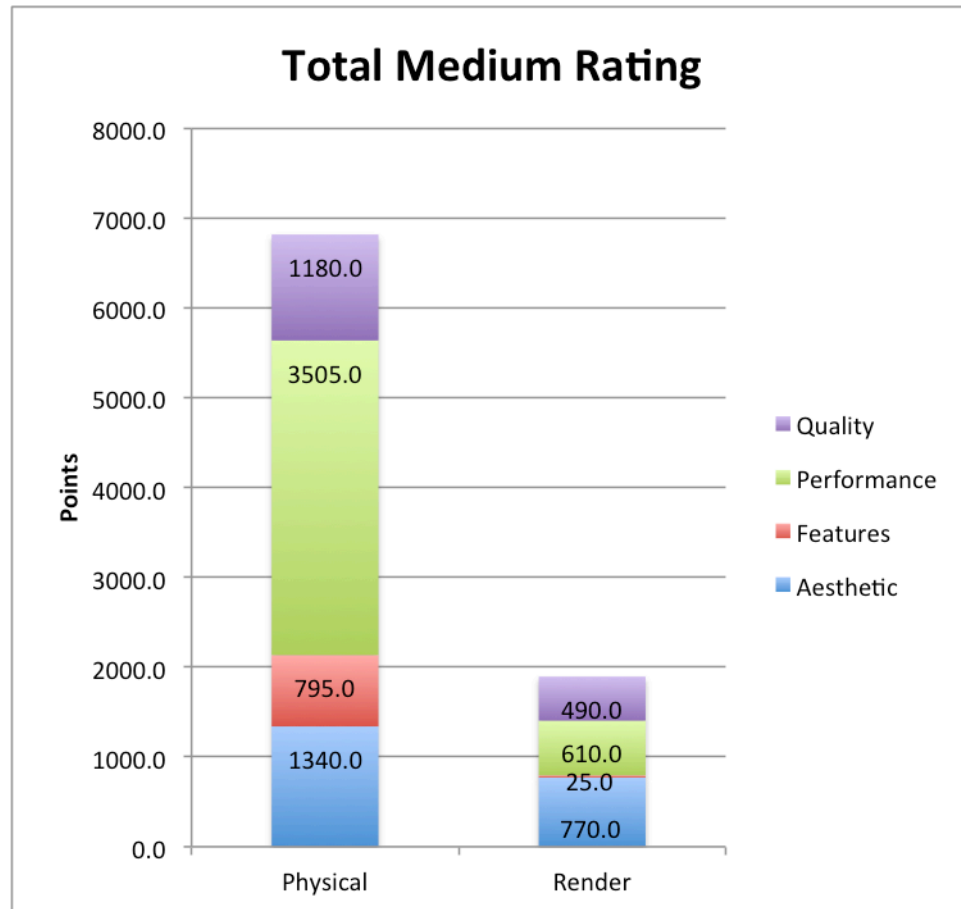


Figure 16. The total rating for each prototype medium. Reported values show the sum of each attribute. Physical prototypes were rated higher.

Differences were found in the amount each attribute contributed to the total rating of prototypes. The percent that each attribute contributed to the sum rating was found, see Figure 17. Figure 16 shows how much each attribute contributed to the total rating of each medium, and Table 2 shows results from two-sample t-tests between attribute percentages. There is a statistically significant difference between the percent that performance and aesthetic ratings contributed to the total rating of each prototype medium. Note that the medium and attribute sum total ratings were lower

for the rendered prototypes than the physical prototypes across the board, and that this analysis only shows each attribute's relative contribution to the total rating.

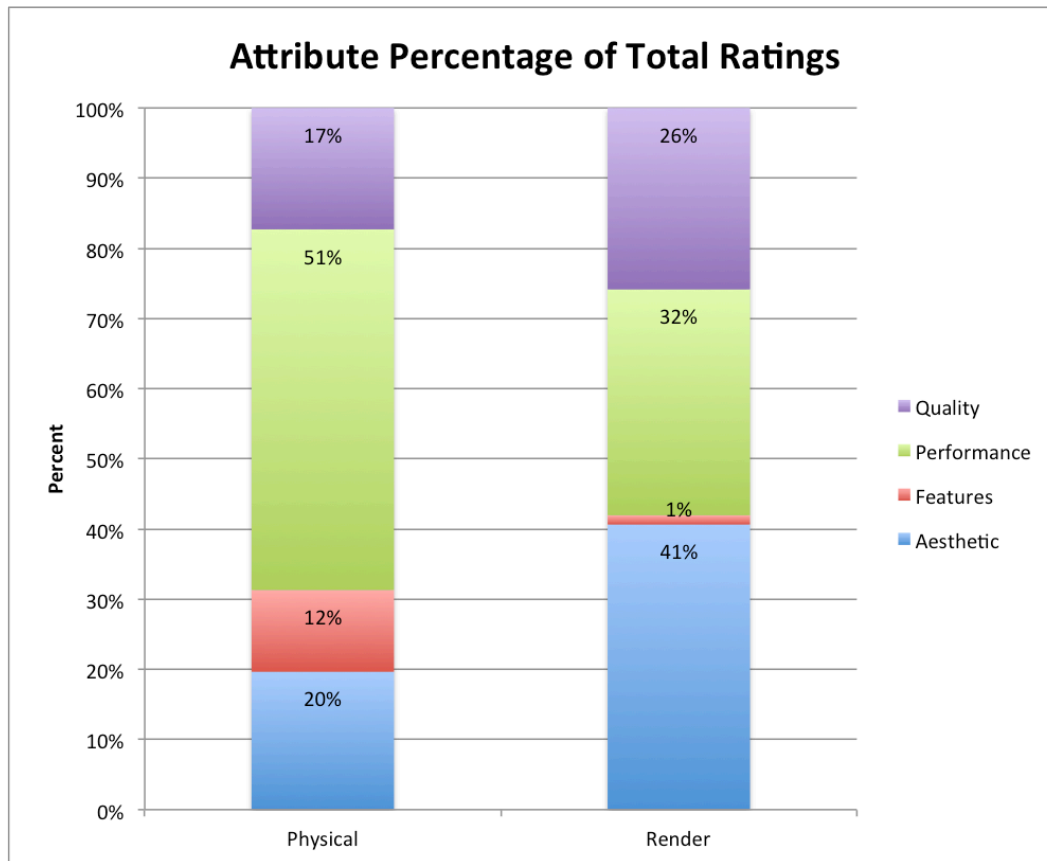


Figure 17. Attribute contribution percentage for total ratings. This figure shows how much points from individual attributes contributed to the total rating of each prototype medium.

Table 2

Summary of attribute contribution differences to total rating between mediums.

*Note significance was $p < .0125$ instead of $p < .05$ because the data was split into four groups.

Attribute	Physical Prototype Percent of Total	Rendered Prototype Percent of Total	Test Statistic	Significance	Significant Difference ($p < .0125$)
Aesthetics	20	41	$t(16)=2.55$	$p=.011$	Yes
Features	12	1	$t(16)=1.76$	$P=.088$	No
Performance	51	31	$t(16)=4.05$	$p < .001$	Yes
Quality	17	26	$t(16)=2.231$	$p=.031$	No

In theory, the attribute contribution should be similar to its average weight, unless the medium of the prototype affects how people evaluate the product. On average the Performance weight was 45 (Figure 14), which is congruent with the Performance contribution (Figure 17) within physical prototypes (51%). However, Performance only made up 31% of the rendered prototypes. Conversely, Aesthetics accounted for a higher percentage of points within rendered prototypes than its average weight. The Features and Quality attributes and the percent contribution distribution were relatively similar throughout. These differences follow the evidence of differences in prototype effectiveness across mediums.

The largest, and only, significant difference in the attribute ratings was from Performance and Aesthetic ratings, which also differed from the weighted expected values in rendered prototypes. These results show that users evaluate rendered prototypes differently, and that a rendered

prototype could more effectively evaluate the aesthetics of a product; however, a physical prototype could more effectively evaluate the performance of a product.

Selected Qualitative Data

The research team wanted to understand how participants came to their conclusions with follow-up questions that asked participants to walk the researchers through their cognitive processes.

Relationship of Valuable Feedback and Prototype Medium

When asked, “Which prototype did you give the most valuable feedback on,” and “In general, did you give more valuable feedback with the physical or rendered prototypes,” all 16 participants reported that they gave the most valuable feedback on the physical prototypes. One participant said it was “easier to evaluate the objects because I can feel them and understood them better.” Another participant said they gave the most valuable feedback on the physical pen because it was “so easy to pick it up and start using it like I was the real thing.” Interestingly, one person said, “It felt more satisfying to rate physical prototypes, like my feedback was more useful.”

Relationship of Prototype and Designer

Responses showed how participants interacted with prototypes within the design process when asked, “If you were a designer which prototype do you think would give you the most valuable information?” One person said, “I think the pen or the door knob. They are both ergonomically complicated and require lots of grip, and precision to use.” Another said, “The objects are easier to integrate the idea of the product and the form.” One participant thought about how she evaluated the Doorknob in rendered and physical form, saying, “The finish for the rendering looks nicer than the 3D printed version, which was kind of distracting because I don’t think of door knobs as rough. But I still don’t like that I can’t interact with the rendering.”

Use of Renderings In The Design Process

Though all participants said they gave more valuable feedback on the physical prototypes, they had insight on renderings in the design process as well. This insight was split between the engineering students and non-engineering students. The former said that renderings could provide valuable feedback in the right scenarios; however, as a whole the latter said renderings were less useful.

One engineering student said that he didn’t like evaluating the rendered prototypes, but he liked how the lines were very crisp and the

product was idealized on the paper. He went on, “It was easy to think about what the product could be instead of with the physical prototypes, which felt more finished.” Another engineering student said, “The renderings were very precise.”

A typical non-engineer response was, “I liked the physical,” “the renderings looked like cartoons,” or “the renders looked fake.” Throughout the feedback process, the non-engineering students gave similar responses that were not as insightful as the engineers.

Population Differences

After talking with participants, it became clear that there was some difference in how participants evaluated prototypes. The qualitative data suggested that people studying engineering evaluate prototypes differently than people not studying engineering. It was not an *a priori* intent of this research to investigate how people with an engineering background evaluate prototypes, but given the stark differences in qualitative responses, the research team investigated potential population differences in the data.

Eight participants were currently studying engineering and design at Tufts University, two were Mechanical Engineering majors, and six were

Human Factors Engineering majors. The other eight participants had neither engineering nor design backgrounds.

A two-way ANOVA was performed comparing the average rating for prototypes between medium and whether the participant studied engineering. No statistically significant differences were found (Table 3). However, there was an interaction between studying engineering and prototype evaluation (Figure 18).

Table 3					
<i>ANOVA of Prototype Medium and Engineering Background Affects on Prototype Effectiveness</i>					
Group	df	Mean Square	F	Significance (p<.05)	Significant ?
Medium	1	13100.788	15.105	p<.001	Yes
Engineer	1	17.721	0.020	p=.886	No
Medium and Engineer	1	512.544	0.591	p=.443	No

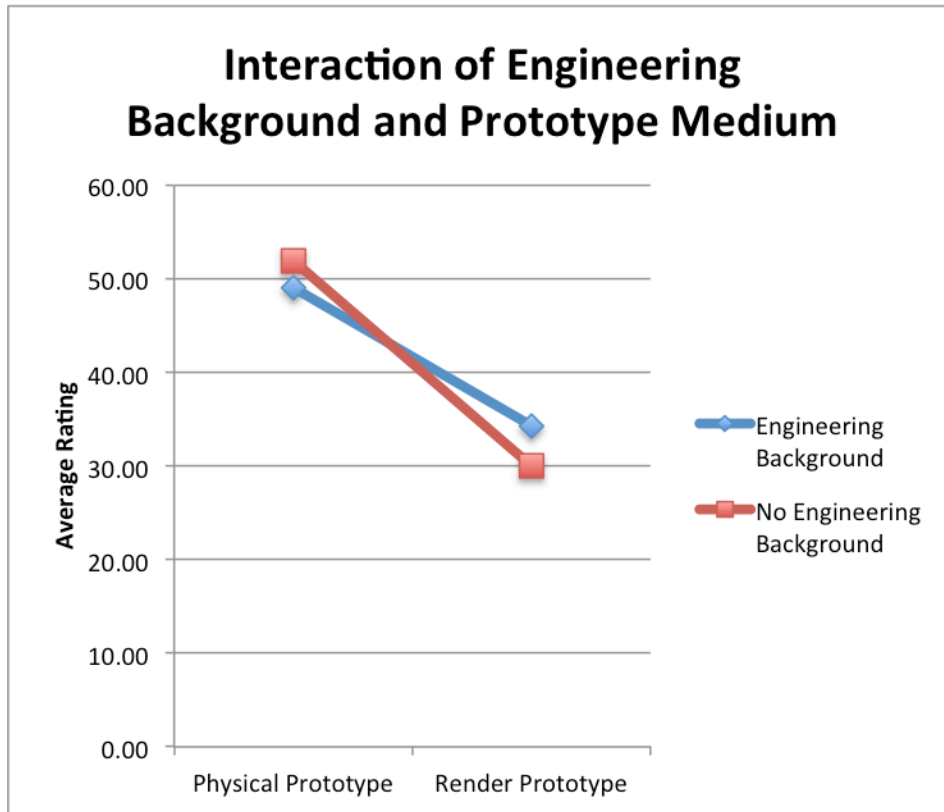


Figure 18. Interaction of Participant Engineering Background and Prototype Medium.

This interaction was found more strongly in qualitative data. Non-engineers responded to the follow-up questions with short answers and minimal description of their thinking as discussed above. In general these answers provided less insight into how participants made their decisions. However, engineering majors rated renderings higher than non-engineers and gave more insight into their decision making process with more verbose answers. In this case insight was how completely they answered the questions. For example an incomplete answer was as simple yes or

no, while a complete answer showed their thought and decision making processes.

A reason for this difference could be that people with an engineering background often extrapolate 3D ideas from 2D descriptions in their studies or work. Previous work has shown that engineers and designers will rate abstract early phase sketches and mock ups as more creative and novel than a layperson (Haggman et al, 2015). Furthermore, Byrne and Johnson-Laird (1989) found that people are generally bad at complex spatial reasoning, which is a highly necessary skill for understanding and interpreting 2D representations of 3D objects. Engineers and designers could be more adept at internalizing 2D concepts and interacting with their mental model of the object.

In practice it is valuable for engineers and designers to target prototypes to the questions they want to answer, and knowing that laypeople understand more about the product when evaluating a physical prototype. Although no significant conclusions can be drawn from population differences in this study, the qualitative data indicates that a difference may exist. If a difference does exist, it is important to note that both participant groups consistently rated physical prototypes as more useful, and all the findings in this research suggest that physical prototypes are more effective in user research.

Relationship Between Product Selection And Primary Study Results

A significant part of this research was product selection. The research team thought that familiar objects would give participants the best chance to fully evaluate and understand the products. Two hypothetical scenarios are proposed. First, a scenario in which unfamiliar products were used, and second, one in which different archetypal products were used.

Alternate Archetypal Products

According to the product selection criteria in the preliminary studies, there were eight options for products that could be used in the primary study. This section analyzes hypothetical results from using a different product in the primary study.

COMPUTER KEYBOARD

In general, computer keyboards garner significant human interaction. There are many usability design concerns such as: tactile feedback, typing ergonomics, material. The research team deemed a keyboard too large to cost effectively 3D print in its entirety, and too complicated to prototype key feedback and interaction.

A keyboard that was 3D printed and painted in the same style as the other physical prototypes would not have working keys, which is a large part of the human interaction. It is hypothesized that participants would fixate on the lack of interaction if it were used in the design feedback portion of the study. For this reason, it is possible that there would be a bias toward rendered prototypes that would skew research results.

LIGHT SWITCH

A light switch is another product with moving parts that requires a certain amount of user interaction. If the Light Switch was used in the study, the physical prototype should have a satisfying click between the on and off positions.

Additionally, light switches are generally placed on a wall and are not free standing objects. Participants could have a hard time evaluating the Light Switch if it was presented with out a wall.

This line of thinking suggests that the Doorknob could have skewed results as well, because doorknobs need to turn to operate, and people normally do not interact with doorknobs that are not attached to doors. However, participants could turn the physical Doorknob in their hands to feel how they might grab it and use it. Therefore, participants would be able to effectively compare the standalone Doorknob. As such, using the

Light Switch in the study may have made a difference in the way users evaluated the physical prototype, but probably not as much as the keyboard.

CREDIT CARD

The research team did not use the Credit Card because it is a standardized form. The only differences between credit cards are the artwork and lettering on them. If it was used in the study, it is highly possible that the rendered prototype would have been rated more useful. As discussed above, renderings are a great medium for evaluating purely visual designs, which is likely what participants would have been done.

DRY ERASE MARKER

Given the similar form and function of the Dry Erase Marker and the Permanent Marker, it is likely the Dry Erase Marker would have performed similarly as the Permanent Marker. This is the same reason the research team decided to only use one of these objects.

Discussion Summary

The results show a significant difference in usefulness across prototype medium. All analysis shows that physical prototypes were considered more effective prototypes in user evaluation. The analyses

examined the product selection process, and differences between mediums, attribute, and object. Qualitative data was used to explain why differences existed, and how decisions made early in the design process affected the results.

CONCLUSION

This research investigated how prototype medium affects user evaluation in the product design process. Key findings related to the main research questions and impacts on the design process are discussed.

Do Physical Prototypes Or Rendered Prototypes Provide A More Effective User Evaluation Of A Product?

Key Finding: Overall participants rated physical prototypes as consistently more effective evaluation of a product.

All the analyses performed in this research indicated that physical prototypes elicited more understanding about a product and more effective feedback. In general, using a physical prototype of a consumer product in a user evaluation study will elicit more useful feedback.

How does the medium of a prototype affect an evaluator's understanding of the prototype on certain design qualities?

Key Finding: Users understand more about a simple consumer product from evaluating a physical prototype.

The qualitative and quantitative results suggested that people understand more about a product when they could interact and manipulate a physical product. People could internalize how physical prototypes affected all of their senses, resulting in a more complete product experience. In practice, design evaluators should realize that physical prototypes can result in more effective user tests because users understand more about the product the prototype represents.

Key Finding: Aesthetics were proportionally better understood in rendered prototypes. Performance was proportionally better understood in physical prototypes.

Aesthetics and Performance ratings were significantly different between physical and rendered prototypes. Rendered prototypes were proportionally more useful for aesthetic evaluation, and less useful for performance evaluation than Physical prototypes. Although rendered prototypes were consistently rated less useful, they still have an important

place in the design process, since they are easy to produce, and effectively evaluate visual aesthetics.

Impacts On The Design Process

Key Finding: Using only rendered prototypes for user evaluation could mask subtle differences between designs because users only partially internalize the product the prototype represents.

For designers to be the most effective in the design process, they should learn as much as possible from every user evaluation. Potentially limiting the amount of information they can learn from a user wastes everyone in the design process' time and money.

In general this research accomplished its goals, with some limitations, as shown with these findings. The results only apply to simple consumer products like those used in this study. There was a limited sample size and future research should use caution before extrapolating these finds to all products and all users. It is very hard to know exactly what someone thinks about a product, and this research used just one of many possible evaluation tools. It is possible that other methods would garner different results.

Finally, it is important to note that with all of the modern prototyping tools there are many ways to solve problems. With all these tools at their disposal, designers need to target their time and energy into using the most effective methods for the question they want to answer. However, the prototype creation *process* can be just as valuable for designers as the prototype itself. As Yang (2005) describes:

The *process* of constructing and refining a 3-D physical prototype can bring up design issues in ways that alternative representations often cannot. While the process of developing of a 2-D drawing or even a computer generated solid model can generate a great deal of information about a design, that information will likely be different than that gained by physically cutting metal.

The usefulness of physical prototypes not only applies to user evaluation. It is also valuable for a designer to create a physical prototype so that they can more completely internalize and understand their own designs.

Future Work

Given the insignificant differences found between participants with engineering and non-engineering backgrounds, future work should examine how a person's background affects prototype effectiveness. This study may have scratched the surface of background differences that

would allow design evaluators to tailor prototypes and tasks to an individual, if they know more about that individual's background. A potential study could use prototypes represented in several mediums, and explore how people's backgrounds affect their understanding of a product. Professional product designers may find greater differences between designs than laypeople. Furthermore, there is potential to understand how much a prototype can be abstracted and still gather effective feedback. The crystallization of the influence of prototype medium on user feedback is a compelling goal that deserves attention in the design research community.

Appendix A

Rate your familiarity with this object



Rating *

1 is low familiarity, 7 is high familiarity

1 2 3 4 5 6 7

Low ☐ ☐ ☐ ☐ ☐ ☐ ☐ High

Name this item *

Figure 19. Sample Preliminary Familiarity Study 1 Question



Figure 20. Archetypal product forms not used in the Primary Study.

Yes/No Priming Questions

1. Do write by hand?	Yes	No
2. Do you use markers?	Yes	No
<hr/>		
1. Do you use markers?	Yes	No
2. Do you draw?	Yes	No
<hr/>		
1. Do you have light fixtures in your house?	Yes	No
2. Do you use a desk lamp?	Yes	No
<hr/>		
1. Have you changed a light bulb?	Yes	No
2. Do you have a floor lamp?	Yes	No
<hr/>		

1. Do you drink water?	Yes	No
2. Do you have glassware in your house?	Yes	No
<hr/>		
1. Do you drink soda?	Yes	No
2. Do you use a reusable cup?	Yes	No
<hr/>		
1. Do you carry keys with you on a daily basis?	Yes	No
2. Do you unlock or lock your door most days?	Yes	No
<hr/>		
1. Does your front door have a deadbolt?	Yes	No
2. Do you have trouble opening your door?	Yes	No

Usability test script
Based on Krug's (2010) script.

Hi, _____. My name is _____, and I'm going to be walking you through this session today.

Before we begin, I have some information for you, and I'm going to read it aloud to make sure that I cover everything.

You probably already have a good idea of why we asked you here, but allow me to go over it again briefly. We are asking people to evaluate and compare several everyday products that we will present to you. We expect this session to take about 60 minutes.

The first thing I want to make clear right away is that we're testing the design of the objects we present to you. You can't do anything wrong here. In fact, this is probably the one place today where you don't have to worry about making mistakes.

Please do not worry about getting the wrong answer. This is entirely subjective and whatever choice you make in each comparison is correct and valid.

If you have any questions as we go along, please ask them and I will answer them to the best of my ability. And if you need to take a break at any point, just let me know.

You may have noticed the microphone. With your permission, we're going to record our conversation as the session progresses. The recording will only be used to help us record your reaction to products during this session. Nobody outside of the research team will ever have access to these recordings. And it helps me, because I don't have to take as many notes.

As for the structure of this session: we will be presenting you with a series of products, please take your time to observe and examine each product that we present to you. Afterwards, we will be asking you to take a brief survey on a computer answering simple questions about your ideas on the products.

Please silence your cell phone so it does not create any distractions.

If you would, I'm going to ask you to sign a simple permission form for us. It just says that we have your permission to record you, and that the recording will only be seen by the people working on the project.

[Give them a consent permission form and a pen]

Design Feedback Session

Now can you fill out this short questionnaire.

[Give them short questionnaire]

Thank you.

Let's get started and get some of your design feedback on this object. I'm going to give you a prototype product. Please give me all your feedback on the design, functionality, and look of the product. I will ask you some questions, but I cannot answer any of your questions for the sake of the study. Please think out loud and say whatever is going on inside your head. Remember there are no wrong answers.

**gives basic product 1 questionnaire*

**Take away questionnaire and pen*

**give participant product prototype 1...2...3...4*

1. Can you give me your first impression of this product?
2. What's a typical use of this product?
3. What do you like about this product?
4. What do you not like about this product?
5. Can you give me your overall impression of this product?
6. Do you have anything else to add? (*Repeat until they don't have anything else*)

Thanks you,

[Repeat for all 4 products]

Opposite Prototype Medium

Participants start taking Survey.

**Set up computer with link preloaded.*

Now I would like to give you an online questionnaire. You can take as much time as you need to complete the survey. Please fill it out completely and thoughtfully.

Please read all instructions and prompts. I'll give you some privacy to complete the survey. If you need anything I will be outside.

Do you have any questions?

**Leave room, go into observation room.*

**Wait for participant to look like they are finished with the task and then reenter the room.*

If it's ok with you I would like to bring in a member of the design team to ask you some questions.

**Andrew comes in and asks clarifying questions from the study.*

Thank you for coming in. Do you have any questions for us?

References

- Artacho-Ramirez, M. A., Diego-Mas, J. A., & Alcaide-Marzal, J. (2008). Influence of the mode of graphical representation on the perception of product aesthetic and emotional features: An exploratory study. *International Journal of Industrial Ergonomics*, 38(11), 942-952.
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data?. *Perspectives On Psychological Science*, 6(1), 3-5.
- Byrne, R. M., & Johnson-Laird, P. N. (1989). Spatial reasoning. *Journal Of Memory And Language*, 28(5), 564-575.
- Crilly, N., Maier, A. M., & Clarkson, P. J. (2008). Representing artefacts as media: Modelling the relationship between designer intent and consumer experience. *International Journal of Design*, 2(3), 15-27.
- Dijk, L., Vergeest, J. S., & Horváth, I. (1998). Testing shape manipulation tools using abstract prototypes. *Design Studies*, 19(2), 187-201.
- Elsen, C., Darses, F., & Leclercq, P. (2011). An anthropo-based standpoint on mediating objects: evolution and extension of industrial design practices. *InDesign Computing and Cognition'10* (pp. 55-74). Springer Netherlands.
- Eppinger, S. D., & Ulrich, K. T. (1995). *Product design and development*. 1995.
- Eugene S. Ferguson. Engineering and the Mind's Eye. *MIT Press*, 1994.

- Fixson, S. K., & Marion, T. J. (2012). Backloading: A Potential Side Effect of Employing Digital Design Tools in New Product Development. *Journal of Product Innovation Management*, 29(S1), 140-156.
- Garvin, D. A. (1984). What does product quality really mean. *Sloan Management Review*, 26(1).
- Gerber, E. (2009). Prototyping: facing uncertainty through small wins. In DS 58-9: Proceedings of ICED 09, the 17th International Conference on Engineering Design, Vol. 9, *Human Behavior in Design*, Palo Alto, CA, USA, 24.-27.08. 2009.
- Häggman, A., Tsai, G., Elsen, C., Honda, T., & Yang, M. C. (2015). Connections between the design tool, design attributes, and user preferences in early stage design. *Journal of Mechanical Design*, 137(7), 071408.
- Houde, S., & Hill, C. (1997). *What do prototypes prototype. Handbook of human-computer interaction*, 2, 367-381.
- Kiefer, S., Silverberg, L., & Gonzalez, M. (2004). A case study of prototyping methods and design for manufacture: electrostatic window blinds. *Journal of Engineering Design*, 15(1), 91-106.
- Krug, S. (2010). *Rocket surgery made easy : the do-it-yourself guide to finding and fixing usability problems*. Berkeley, Calif: New Riders.

- Kudrowitz, B. M., & Wallace, D. (2013). Assessing the quality of ideas from prolific, early-stage product ideation. *Journal of Engineering Design*, 24(2), 120-139.
- Macomber, B., & Yang, M. (2011, January). The role of sketch finish and style in user responses to early stage design concepts. In *ASME 2011 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (pp. 567-576). American Society of Mechanical Engineers.
- Martin, R. C. (2003). *Agile software development: principles, patterns, and practices*. Prentice Hall PTR.
- Norman, D. (2004). *Emotional design : why we love (or hate) everyday things*. New York: Basic Books.
- Reid, T. N., MacDonald, E. F., & Du, P. (2013). Impact of Product Design Representation on Customer Judgment. *Journal of Mechanical Design*, 135(9), 091008.
- Rosner, B. (2011). *Fundamentals of biostatistics*. Boston: Brooks/Cole, Cengage Learning.
- Rubin, J. & Chisnell, D. (2008). *Handbook of usability testing how to plan, design, and conduct effective tests*. Indianapolis, IN: Wiley Pub.
- Schrage, M. (1993). The culture (s) of prototyping. *Design Management Journal* (Former Series), 4(1), 55-65.
- Schumann, J., Strothotte, T., Laser, S., & Raab, A. (1996, April). Assessing the effect of non-photorealistic rendered images in CAD.

In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 35-41). ACM.

Snyder, C. (2003). *Paper prototyping the fast and easy way to design and refine user interfaces*. San Francisco, Calif: Morgan Kaufmann, Elsevier Science.

Söderman, M. (2005). Virtual reality in product evaluations with potential customers: An exploratory study comparing virtual reality with conventional product representations. *Journal Of Engineering Design*, 16(3), 311-328.

Ullman, D. G., Wood, S., & Craig, D. (1990). The importance of drawing in the mechanical design process. *Computers & Graphics*, 14(2), 263-274.

Ullman, David G. *The mechanical design process*. Vol. 2. New York: McGraw-Hill, 1992.

Ulrich, L. (2015). *Tangible things : making history through objects*. New York: Oxford University Press.

Vest, C. M. (2005). *Educating engineers for 2020 and beyond*. National Academy of Engineering.

Wickens, C., Gordon, S. & Liu, Y. (2004). *An introduction to human factors engineering*. Upper Saddle River, N.J: Pearson Prentice Hall.

Yang, M. C. (2005). A study of prototypes, design activity, and design outcome. *Design Studies*, 26(6), 649-669.