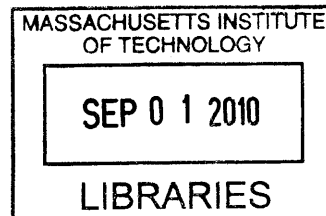# HaHa and Aha!

Creativity, Idea Generation, Improvisational Humor, and Product Design

by

Barry Matthew Kudrowitz

M.S. Mechanical Engineering
MIT, 2006

S.B Mechanical Engineering
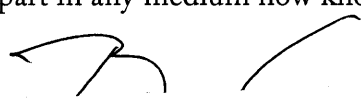University of Central Florida, 2004

Submitted to the Department of Mechanical Engineering on May 19, 2010 in Partial
Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Mechanical
Engineering at the Massachusetts Institute of Technology

JUNE 2010

Signature of Author _____ _____

Barry Matthew Kudrowitz
Department of Mechanical Engineering
May 19, 2010

Certified by _____

David Wallace
Professor of Mechanical Engineering
Thesis Supervisor

Accepted by_____

David E. Hardt
Professor of Mechanical Engineering
Graduate Chair

HaHa and Aha!

Creativity, Idea Generation, Improvisational Humor, and Product Design

by

Barry Matthew Kudrowitz

Submitted to the Department of Mechanical Engineering on May 19, 2010 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Mechanical Engineering

Abstract

It is widely recognized that innovation and creativity is the new competitive battleground for product development firms. Engineers and product designers are now expected to be highly creative, prolific idea generators in addition to being analytically competent. Thus, it is of interest to study methods to improve a designer's idea generation capabilities. It is believed that wit, being spontaneous humor production, is strongly related to creativity as both involve making non-obvious connections between seemingly unrelated things. This thesis looks into the realm of humor and improvisational comedy to suggest means of enhancing creative output in blue-sky product design idea generation.

We have found that the ability to quickly generate many ideas is strongly correlated ($r^2=.82$) with being able to come up with a single, promising, creative idea. It was also found that, with appropriate training, individuals may learn to become more prolific idea generators. Furthermore, improvisational comedians were more proficient at new product idea generation than professional product designers, and methods for training comedians can be effectively adapted to product design idea generation.

In a study where 84 participants (students, professional designers and improvisational comedians) took a cartoon caption humor test and a nominal product brainstorming test, we found that improvisational comedians on average produced 20% more product ideas and 25% more creative product ideas than professional product designers. Furthermore, the few individuals that were highly prolific in both creative product ideation and humorous cartoon caption production had an improvisational comedy background. Many of the games used in improvisational comedy training are intended to promote associative thinking. We designed an improvisational comedy workshop composed of these association-based games. A group of 11 subjects who participated in this workshop increased their idea output on average by 37% in a subsequent product brainstorming session. Our findings suggest that improvisational comedy games are a useful warm-up for idea generation, that prolific generation is not a domain-specific ability and that it is possible to teach creativity. Ultimately, this work can lead to the development of tools and methods that designers can use to improve their idea generation skills.

Thesis Supervisor: David Wallace
Title: Professor of Mechanical Engineering

Acknowledgements

Chapters

8

List of Figures

List of Tables

1. Introduction

```annd wee mussttn'tt looose ourr sensses of hummorr,' Mrs. Which said. `The onnlly wway ttoo ccope withh ssometthingg ddeadly sseriouss iss ttoo ttry ttoo trreatt itt a llittlle lligghtly.[1]"
–Madeline L'Engle, A Wrinkle in Time

Creativity is the new core competence in engineering and design. "It isn't just about math and science anymore. It's about creativity, imagination, and, above all, innovation [2]." It is believed that the basis of creativity is the ability to make non-obvious connections between seemingly unrelated things [3-5]. This is also believed to be the basis of humor or more specifically wittiness [5-8]. Comically, this thesis is self-referencing—while the thesis is about making non-obvious connections, it makes the non-obvious connections between creativity and humor.

There are prior correlation studies that compare facets of humor with creativity tests, but none address the application to design practice and industry [9-15]. Contemporary research has suggested that designers can learn from improvisational comedy, but there is no empirical data to support these claims [16, 17].

In this study, we found that being able to generate many ideas is highly correlated (r^2=.82) with being able to come up with promising creative ideas as shown in Figure 1. We also found that improvisational comedians produce 20% more product ideas and 25% more creative ideas than professional product designers. Furthermore, we found that improvisational training can increase idea output on average by 37% in a subsequent product brainstorming.
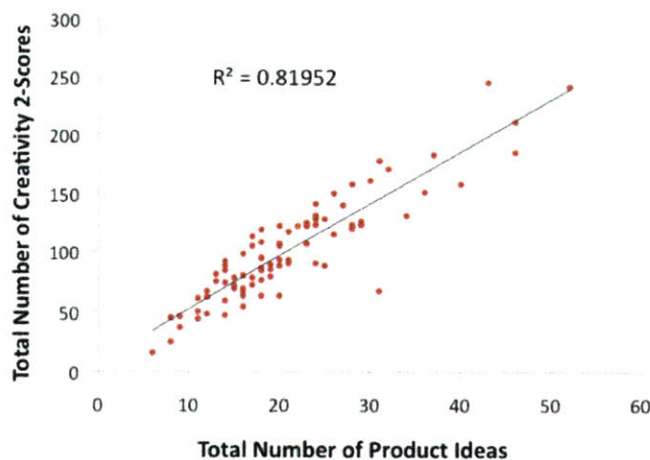
Figure 1. Overall Quantity of Ideas vs. Creativity Score per Subject

13

The goal of this work is to look into the realm of humor, and more specifically, improvisational comedy to find new ways of improving upon the idea generation process to make designers and engineers more prolific, creative idea generators.

This thesis:

1) Provides empirical data supporting the relationships between humor production ability and creativity in product design ideation,

2) Builds on the research in testing humor production ability,

3) Explores the use of nominal product brainstorming as a measure of applied creativity,

4) Explores differences in humor production and product ideation between professional designers, students, and improvisational comedians,

5) Determines what qualities people associate with innovative product ideas,

6) Explores how quantity correlates with quality in both humor and idea production,

7) Explores how presentation (wording of humor and sketching of ideas) influences appreciation of the content, and

8) Provides empirical data supporting the effectiveness of improvisational comedy training on enhancing ideation fluency

2. Prior Art

2.1 Wit: The Connection between Humor and Creativity

"the person who is spontaneously humorous is, by the same token, spontaneously creative [6]."
~ J.D. Goodchilds

There have been several researchers and philosophers that have theorized connections between humor and creativity [3, 5-8]. There have also been several experimental studies that have specifically dealt with correlating humor and creativity [9-15] and most have found moderately positive correlations. Many studies attempt to correlate general humor with general creativity; however, "not every creative solution is humorous, and many instances of humor are far from creative [11]." Wit is where the two realms overlap most clearly [8].

Wit is a subset of humor that describes intentional, spontaneous humor production. The following quote summarizes the key differences between wit and humor:

> ...Wit is always intentional, humour is always unintentional... [w]it depends for its success upon condensation, revelation, suddenness and surprise, and it necessitates a quick and deliberate motion of the mind; it is not private indulgence but invariably needs an audience, it is thus a social phenomenon. [18].

The Oxford American Dictionary distinguishes between humor and wit as humor relying less on intelligence [19].

As opposed to humor which is viewed as an inborn personality trait, wit is considered to be more artificial and something that could be acquired through learning and practice [20, 21]. Similarly, creative thinking skills can be acquired through learning and practice [22]. It is the goal of this work to explore ways of making engineers and designers more witty and creative.

2.1.1 Making Non-Obvious Connections Between Seemingly Unrelated Things

"The world is full of obvious things which nobody by any chance ever observes"
- Arthur Conan Doyle, Sr

When comparing the manners in which researchers and theorists describe wit and creativity, there is a clear overlap in definitions. Both wit and creativity connect previously unrelated dimensions of experience [13].

"The creative act of the humorist consists in bringing about a momentary fusion between two habitually incompatible frames of reference" [5]. In accordance with the Incongruity Theory of Humor, something is funny as a result of a pairing of ideas or situations that are divergent from expected [23].

The creative act, whether in poetry or science, depends on discovering analogies between two or more ideas previously thought unrelated [4]. In accordance with the Associative Theory of Creativity, creativity is "the forming of associative elements into new combinations which either meet specified requirements or are in some way useful [3]."

It seems as if the Associative Theory of Creativity [3] greatly overlaps with the Incongruity Theory of Humor [23].

In addition, both theories explain the degree of creativity or humor in terms of remoteness of association. What makes a joke funny is the amount that the punch line violates the recipient's expectations while assuming the recipient is still able to make sense of the information [24]. Putting this in terms of creativity, "the more mutually remote the elements of the new combination, the more creative the process or solution [3]." When the connections are too obvious, the output is not considered creative or funny. When connections cannot be made, then the output is considered confusing.

2.1.2 Prior Humor and Creativity Correlation Studies

There are several studies that have attempted to correlate a sense of humor with creative abilities and the results are mostly positive as shown in Table 1. Many of these studies correlate a general sense of humor with general creative abilities. One study chose to use the Multidimensional Sense

16

of Humor Scale because it taps into many different types of humor including appreciation, playfulness, and coping ability [15]. We are interested, however, in wit or humor production as we believe this is where the two realms overlap most clearly [8]. In the few studies that have dealt with creation of humor (bolded in the table) and not simply humor appreciation, three involved the participants adding a funny caption to a caption-stripped cartoon [10, 11, 13]. The two more contemporary studies did not provide any details on the humor production test [14, 25].

As one can see in Table 1, most of these correlation studies have positive r^2 values indicating low-moderate correlation. It is possible however that a third factor, such as intelligence or verbal fluency, is a hidden variable that connects humor and creativity.

Table 1. Prior Correlation Studies of Humor and Creativity (Humor Production Tests in Bold)

Study	Test of Humor	Test of Creativity	Maximum R Value
1965 Smith and White [9]	Self Survey, Peer Rating	Word Association	+.18 (p<.05)
1970 Treadwell [10]	**Cartoon caption**, Self Report	Remote Associates Test, Gestalt Transformation	+.275 (p=.05)
1974 Babad [11]	**Cartoon caption**, humor appreciation, peer rating	Torrance's Circles, Sentences	+.08 (p<.01)
1975 Rouff [12]	Explaining Cartoon Humor	Remote Associates Test	+.37 (p<.001)
1976 Brodzinsky and Rubien [13]	**Cartoon caption**	Remote Associates Test	positive
1980 Clabby [14]	**Campaign slogan creation**	Alternative uses	+.325 (p<.01)
1996 Humke and Scheafer [15]	Multidimensional Sense of Humor Scale	Franck Drawing Completion	+.77 (p=.01)
2004 Sitton and Pierce [25]	**Geographic Pun Generation**	Remote Associates Test	+.42 (p=.001)

The studies in Table 1 use a variety of creativity tests, but none of these examine applied creativity, suggest tools for increasing witty/creative thought, or compare different demographics.

2.2 Innovation, Invention and Creativity

"If at first, the idea is not absurd, then there is no hope for it."
~Albert Einstein

Creativity is the ability to produce work that is both novel (i.e. original, unexpected) and appropriate (i.e. adaptive/appropriate to the problem, of value) [26, 27]. This definition is inline with what is called a "product definition of creativity" where creativity is defined and measured by the output of a process and not by the person or the process itself [28]. Although a universally

precise applicable definition of creativity does not exist [28], the product definition of creativity is appropriate for the scope of this study

When creativity is applied to the realms of design, engineering or science, it is sometimes referred to as innovation or invention.

Invention is creating a new and useful technology. *Innovation* is the combination of knowledge or technologies in original and non-obvious valued new products, processes or services [29]. "A discovery that goes no further than the laboratory remains an invention" [30].

All inventions and innovations are creative, but not all creations are innovative or inventive. All innovations start from inventions, but not all inventions become innovations.

2.2.1 Classifications of Novelty and Innovation

"Classification is the first step in bringing order into any scientific endeavor"
~Michael Ashby and Kara Johnson [31]

Novelty, and therefore creativity and innovation, are culture or society dependent. There is said to exist three types of novelty: novel to the individual, novel to the society/group, novel to history [32]. In the realm of product design, societal and historical novelty is more important than individual novelty. An idea for a product that is only novel to the individual would most likely not be considered innovative.

Innovation, in the form of products, can exist at a micro and macro level. At a macro level, a product innovation is new to the world, the market or the industry. At a micro level a product innovation is new to the customer or the firm [30]. There are also three classes of innovation: Radical Innovation, Really New Innovation, and Incremental Innovation. Radical innovation is both at a macro and micro level and embodies a new technology that creates a new market. Really New Innovations are moderate innovations between incremental and radical that create new lines of products or new markets with existing technology. Incremental innovations are micro-level innovations that provide new features, benefits or improvements to existing technology in an existing market [30].

Incremental innovations are easier to be understood and accepted by society than radical innovations as they can be easily assimilated into existing mental and societal framework [33].

18

People prefer designs that are novel as long as the novelty does not affect the typicality of the design [34, 35]. This concept is sometimes referred to as MAYA or most advanced yet acceptable [34]. The shift from an old model of a car to a newer more modern model of car (e.g. the VW Beetle) is an example of an incremental innovation that is both novel and similar enough to the established image. The preference for novelty combined with typicality is present in more than just products and artwork, but also music and food. This concept could explain the popularity of parody and sampling in the music and film.

2.2.2 Special Classes of Innovation

Innovation "is the whim of an elite before it becomes a need of the public. The luxury today is the necessity of tomorrow."
~ Ludwig Von Mises

There are several groups of concepts and products that are in the grey area of the innovative spectrum. Many definitions of creative and innovation include the element of "useful," "practical," or "appropriate" as a key feature. Sometimes concepts are considered innovative or creative without having this "useful" element [36, 37]. This group of "un-usefully" novel concepts and products goes by several different names and comes in several different forms.

2.2.2.1 Aesthetic Innovation

An *aesthetic innovation* is the transformation and manipulation of a product's appearance, including changes made to the materials, colors, proportions, textures, shape, or ornamentation [38]. In other words, aesthetic innovation is novel changes in a concept that do not involve the use function; it is a form of incremental innovation. A toaster that is shaped like a loaf of bread or a toaster that is completely clear would be examples of aesthetic innovation in the toaster product realm. However, if the product's main function were aesthetics (i.e. art), than the term aesthetic innovation would be redundant.

2.2.2.2 Novelty Items and Gadgets

Novelty items are products that people buy that have no practical function. There are stores that are dedicated to items that are only marginally useful such as The Sharper Image and Spencer's. A more contemporary product example is the Juicy Salif designed by Philippe Starck for Alessi in 1990. This is a juicer that is not intended to be used as a juicer, but more for decoration and conversation. According to Donald Norman the designer allegedly said that his "juicer is not meant to squeeze lemons; it is meant to start conversations[39]." The gold plated version of the

19

juicer is intended to be an ornament because the citric acid in a lemon discolors and erodes the gold plating. Gadgets and gizmos are the names typically given to novelty items when they have electronic components and perform tasks that are superfluous.

Toys are sometimes considered novelty items, as they are not "useful" or "practical." However, play is considered to be the function of a toy product, and therefore items that are designed solely for play, should still be considered useful [40]. However, toys, like the Pet Rock and the Big Mouth Billy Bass, are considered novelty items when they do not have play value. Designer toys are collectables produced in limited editions and are in essence action figures for adults. These toys are typically not played with and are treated more like works of art and means of self-expression [40]. Designer toys could be considered novelty items.

2.2.2.3 Chindogu and "Unuseless"

Chindogu is the Japanese art of creating products that seem to solve a problem, but actually cause new problems or social embarrassment to the point that they have no utility. [41]. The term *Chindogu*, which means "unusual" was coined by Kenji Kawakami and was made popular by his book called "The Big Bento Box of Unuseless Japanese Inventions" [42]. These items are sometimes referred to as "unuseless" as they are essentially useful, but extremely impractical to the point that they are no longer useful. A few examples of Chindogu are: a hay fever hat which is actually a toilet paper roll fixed to the top of a hat and shoulder pad baskets to catch earrings if they should fall out.

The products that fall into this "un-useful" category are similar to nonsense humor where the novelty is not quite justified with utility, however, the gestalt is taken as innovative or at least product worthy.

2.3 Creativity in the Product Design Process

"Anyone who can be replaced by a machine deserves to be."
~Dennis Gunton

Creativity is a critical part of the design process. Without creative design, culture and society may remain stagnant. Design problems do not have one correct solution and often there are many solutions for any given problem. It takes creativity on the part of the designer to address the problems in new ways to develop novel solutions. It is the creative element that is the less

common, less taught, less understood, yet more desired and influential aspect of design. The immediate importance of creativity in the design industry is perhaps best described by this following quote by Bruce Nussbaum in BusinessWeek:

> What was once central to corporations -- price, quality, and much of the left-brain, digitized analytical work associated with knowledge -- is fast being shipped off to lower-paid, highly trained Chinese and Indians, as well as Hungarians, Czechs, and Russians. Increasingly, the new core competence is creativity -- the right-brain stuff that smart companies are now harnessing to generate top-line growth. The game is changing. It isn't just about math and science anymore. It's about creativity, imagination, and, above all, innovation [2].

A product design process can be described as a series of stages. Ulrich and Eppinger describe six phases: Planning, Concept Development, System-Level Design, Detail Design, Testing and Refinement, and Production Ramp-up [43]. Niku visualizes the process in a more iterative manner where the designer goes through a series of stages and may have to revisit the stages at different points in the process. Niku has five stages and focuses more on developing the ideas than developing the artifact: initiation (need finding, understanding the state of the art), specification (development of the problem and requirements), ideation, implementation (including detailed design and refinement), and iteration [44]. Creativity is important in all stages of design, but at the beginning, during idea generation, is where creativity can have the most impact. A bad idea implemented well is still a bad idea.

2.3.1 Idea Generation Methods

"Think left and think right and think low and think high. Oh, the thinks you can think up if only you try! [45]"
~Dr. Seuss, Oh, The Thinks You Can Think!

There are many methods that designers use to come up with new ideas for products and creative solutions to problems. These idea generation (or *ideation*) methods can be grouped into categories in several ways.

Smith [46] studied 172 idea-generation methods and identified the active ingredients, or devices, in each method that promote the idea generation. Smith found 50 different devices that he classifies into three types (strategies, tactics, and enablers). Strategies are the most common and

are mental activities and procedures (decomposition, association, change perspective, combination, group interaction, etc). Tactics are stimulatory tools that support the strategies (display ideas, change environments, elaborate, etc). Enablers are passive means of generating ideas (block removal, goal setting, mass production, incubation). Tactics and enablers are not methods in themselves, but tools and techniques that are used with methods.

Another way of classifying methods is by the type of problems they are addressing. Problems can be ill structured (free form) when they have poorly defined goals [47] such as "we need new forms of water toys." These types of problems often have a large number of solutions and there is no correct or best answer. Problems that are structured, such as "find the least expensive means of securely hanging this picture frame," have well defined goals and often there is an optimal solution.

Just as there are ill structured (or freeform) and structured problems, there are also freeform and structured idea generation methods. Freeform idea generation methods include: brainstorming, free association and brainwriting; and structured idea generation methods include TRIZ, morphological analysis, and SCAMPER [43, 48]. The freeform methods tend to favor blue-sky design and initial stages of design (typically performed as a group), while the structured methods are perhaps more suitable for detailed design stages (typically performed by individuals).

The authors' design pedagogy and research practices use mostly free form ideation methods and, in particular, the process known as brainstorming.

2.3.2 Play in Idea Generation

"If you want creative workers, give them enough time to play"
~John Cleese

There are several books that attempt to teach creativity by explaining ways of thinking differently. A Whack on the Side of the Head by Roger Van Oech is series of mind games to inspire creativity [49]. Thinkertoys by Michael Michalko is a collection of over thirty tools and techniques in the style of games to inspire creativity thinking [48]. Edward De Bono wrote several books (Serious Creativity, Six Thinking Hats, Lateral Thinking) on ways of thinking differently [50-52].

All of these books treat creativity and idea generation as play. Playfulness has been found to correlate with divergent thinking [53], associative fluency [54, 55] and higher scores on creative

tests [56]. Simply having fun and being in a state of positive affect arouses curiosity, reduces anxiety and engages creativity [39] and was shown to increase creative thought processes [57-59].

2.3.3 Association Making in Idea Generation

"There is only one way in which a person acquires a new idea: by the combination or association of two or more ideas he already has into a new juxtaposition in such a manner as to discover a relationship among them of which he was not previously aware."
~Francis A. Cartier

Making connections or associations is an important component of many idea generation methods (Brute Think, Analogies, Mind Mapping, Provocation [48]). More than half (26/50) of the idea generation devices proposed by Smith [46] explicitly involve making some form of association.

MacCrimmon and Wagner, describe four ways in which connections can be made between existing ideas: Internally (either with form and function or purpose) or externally (either locally or distant) [60]. Making connections with the form and function involve the apparent attributes (shape, material, weight, mechanisms). Making connections with the purpose involves the goals (how is something used, where is it used, why is it used). Making connections with a local source involves mapping a related problem/product to the current problem/product. Finally, making connections with a distant source involves attempting to map something that is seemingly unrelated to the problem/product.

Making connections with distant sources is similar to the idea generation tool known as brute think. Michalko describes brute think as forcing a connection between two dissimilar concepts to create a new idea [48]. When using brute think in an idea generation session, a random word is chosen from a list, the dictionary, or another book and participants attempt to incorporate the word into the problem at hand.

A tool called Association Mapping [61] can help in the process of making non-obvious associations. This tool, as shown in Figure 2, is based on free association and mind mapping [62] where a concept or item is placed in the center of a page and attributes related to that entity are written around it. Words that describe each attribute are noted and then further associations are made from those words. This method is a bit more structured than free association [48] and it adds the element of mapping back to the original domain, which makes it both a divergent and convergent process. Associative thinking can help a designer address a problem in non-obvious manners. It may also have potential to make people be wittier.
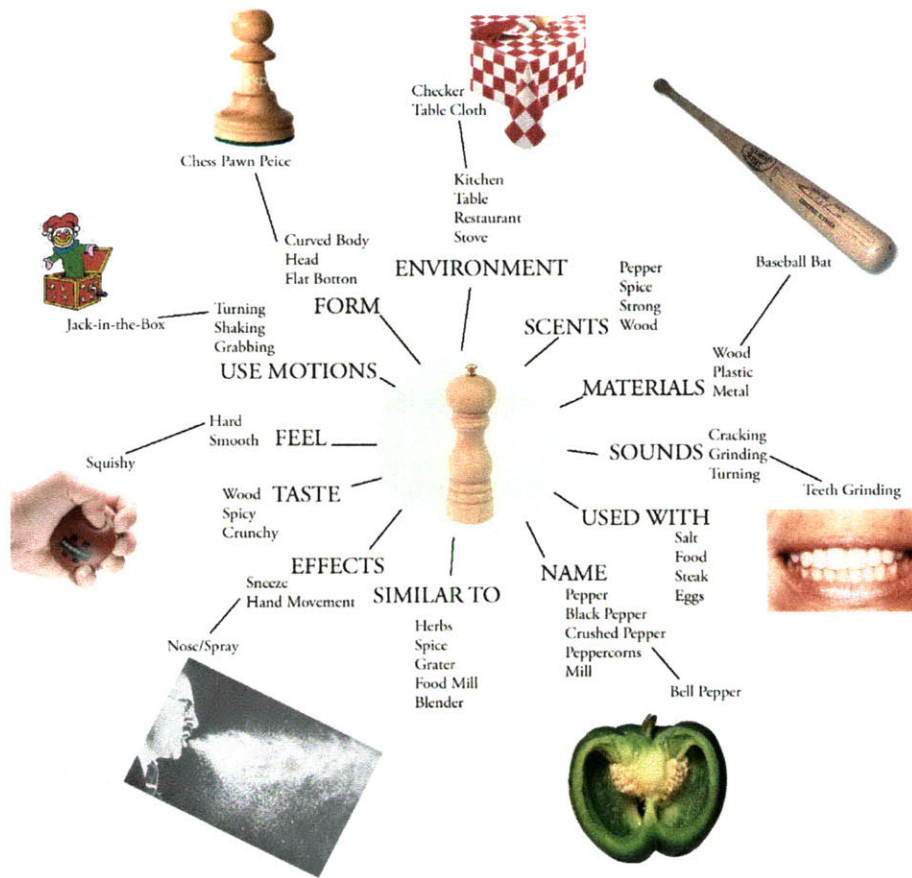
Figure 2. An Association Map Example for a Pepper Mill

2.3.4 Brainstorming

Brainstorming is the most common method for generating conceptual designs. Brainstorming sessions are critical for building consensus, communicating information, and refining concepts.

The formal concept of brainstorming was developed by Alex Osborn in the 1930s and written about in his book Applied Imagination [63]. In a brainstorm session, a group sits together to generate a large number of ideas in a short amount of time. The participants quickly sketch any idea that comes to mind and present it to the group. Each sketch should be drawn large and clearly on a separate sheet of paper and given an appropriate and simple title. As ideas are generated, a group facilitator calls on members to share their ideas. The group facilitator ensures everyone hears each idea and then records the concept title with a brief description.

The four main rules of a good brainstorm session are to defer judgment and not critique ideas, build off of each other's ideas, encourage wild ideas and come up with as many ideas as possible [63]. All participants must be able to speak freely without fear of being subjected to harassment. No ideas are to be considered "silly" or discarded as infeasible, as these irrelevant or fanciful ideas may lead to plausible concepts.

Brainstorming groups should be composed of a minimum of four members to insure a significant amount of views, opinions, and therefore, concepts [44]. The optimum size is about a dozen [63]. A group should also be limited to around a dozen members to avoid segmentation within the group. All ideas should be recorded for later review and legal documentation. Finally, sessions should not last longer than 30 to 45 minutes, as people tend to become less productive [44, 63].

2.3.5 Nominal Brainstorming

In a nominal brainstorm session, the participants do not talk to each other. They either share their ideas silently or do not share the ideas until after the session ends. Nominal brainstorms are used in studies when individual ideation ability is of interest. It has also been argued that nominal brainstorming is more effective than traditional brainstorming as it eliminates three theoretical factors of productivity loss: production blocking, evaluation apprehension and free riding [64]. Production blocking is the term given to the phenomenon that occurs when participants in a brainstorm session suppress or forget their ideas because they seem less relevant or other speaking members distract them. Evaluation apprehension is the term for the fear of negative evaluation from other group members. Free riding is the term for participants not contributing to the group as they feel that their contributions are dispensable. As the number of group members increases, the number of free riders also increases [65].

Although nominal brainstorming may have shown to be more productive than traditional brainstorming in several studies [64, 66], there are several benefits of the traditional interactive brainstorming session. In an interactive brainstorm session, participants can easily build off of each other's ideas. Participants can become more emotional and expressive with their ideas. With interaction, these sessions can be used for group bonding in addition to ideation. Participants can also become more comfortable presenting their ideas to each other.

2.4 Measuring Creative Abilities

They say, "Come up with somethin' new"
And everyone will buy it."
So I came up with a paper umbrella.
But no one was willing to try it.
And then I came up with reusuable gum.
It seemed such a pity to waste it.
Then I came up with some mustard ice cream.
Nobody bothered to taste it.
So now I've invented a plug-bottom boat.
It's just what you need, there's no doubt.
'Cause if any water should ever splash in.
Just pull the plug—it'll all run out. [sic] [67]
From Shel Silverstein's "Something New"

If we are intending to compare individual wit and creativity we need to be able to measure creative abilities. There are two basic types of creativity studies: archival and experimental [68]. Archival studies are ones that look at the social variables that influence personality based on information from biographies, histories, etc. These studies are rare, ultimately require experimental studies to support any findings, and depend on extensive documented materials of individuals' histories. Experimental studies can be divided into two categories: *creativity tests*, which are instruments administered like intelligence tests, and *product ratings*, where judges assess the creativity of products. The word "products" is used in the general sense as meaning output, rather than the specific reference to items intended for manufacture [68].

2.4.1 Creativity Tests

The majority of experimental creativity studies are based on creativity tests as they are relatively simple to administer and the data can be analyzed objectively. A few of the well known creativity tests include the Torrance Test of Creative Thinking (TTCT) by Paul Torrance [69], the Remote Associates Test (RAT) by Sarnoff Mednick [3] and the Guilford's Structure of the Intellect (SOI) divergent production tests by Joy Paul Guilford [70]. These tests measure certain cognitive abilities that have been shown to correlate with creative thought processes such as divergent thinking.

2.4.1.1 Remote Associates Test (RAT) - 1962

The Remote Associates Test (RAT) was developed along with the associative theory of creativity, which claims that creative thinking involves forming of associative elements into new combinations

that meet specified requirements or are in some way useful [3]. The test involves finding a connective link between sets of three seemingly unrelated words that have a mutually remote association. An example of a set of words could be: tap rain floor. The subject would be required to find the word that serves as a connective link, which can be paired with any of the three words in the set. For the given example, the word "dance" connects with the three words in the forms of "tap dance," "rain dance" and "dance floor."

The RAT was found to correlate negatively with GPA for undergrads [3], correlate positively with brainstorming productivity [71], and correlate more positively with convergent thinking tests rather than divergent thinking tests [72]. The RAT was also found to positively correlate with humor comprehension [12], and humor production [10, 13, 25]. As making sense of a joke and creating a joke involves making a connection between seemingly unrelated words, it seems reasonable that the RAT correlates with both humor comprehension and creation.

2.4.1.2 The Torrance Test of Creative Thinking (TTCT) - 1966

The Torrance Test of Creative Thinking (TTCT) is a series of creativity tests that involve oral, written or drawn responses. Derived from Guilford [70], these tests are scored on four criterion components of creativity: 1) Fluency, the production of large numbers of ideas; 2) flexibility, the production of a large variety of ideas; 3) elaboration, the development, embellishment or filling out of ideas; and 4) originality, the use of ideas that are not obvious or statistically frequent [28].

One of the more popular tests in the TTCT is the "Circles Test" where the participant is asked to sketch as many different objects as possible using a set of blank circles and provide a title for each sketch. Fluency is the count of the number of circles used, flexibility is the number of different categories or themes of objects, originality is scored in points for objects that are statistically infrequent in the population of test subjects, and elaboration is scored by the number of details used in the sketches.

Some other tests of the TTCT include: Asking questions- writing down questions based on a drawing; Product improvement- listing ways to change a toy monkey for greater play; and Unusual Uses – based on Guilford's test, list unusual uses for a cardboard box. [26]

2.4.1.3 Guilford's Structure of Intellect (SOI) – 1956

Guilford's Structure of Intellect (SOI) is the basis of many other creativity tests including the TTCT [28]. Guilford was the first to propose that it is possible to study and evaluate creativity of subjects using a psychometric approach with pencil and paper [26].

The "Unusual Uses Test" test is perhaps the most cited of his tests and it appears in other creativity tests sometimes called "Alternative Uses." In this test, the subject is asked to think of as many unconventional uses for each of a number of objects, the number given being the score [70]. The objects that are typically used include: a brick, a newspaper, and a cardboard box. This test is a measure of "spontaneous flexibility" according to Guilford [70], which is one of the four criterion components adapted by Torrance: fluency, flexibility, originality and elaboration. Guilford's tests began the usage of divergent thinking tests as the main instrument for measuring creativity [26]. The Unusual Uses or Alternative Uses creativity tests are perhaps the most similar to a product idea generation session.

2.4.2 Product Rating Tests

In this study, we focus on product rating for an assessment of the participants' creative abilities. There are several reasons why a product rating, as opposed to a standardized creativity test, is best suited for this work. In the product design industry, what ultimately matters is the product. If an individual produces creative products ideas, it is less important how they score on a standardized creativity test. As Amabile states "any identification of a thought process as creative must finally depend on the fruit of that process—a product or response. …Not only does the task [product creation] itself mimic real-world performance, but the assessment technique mimics real-world evaluations of creative work. [68]." O'Quin and Besemer agree with the view that judgments of products are more stable and more valid than the standardized tests of creativity [73].

The word *product* refers to both artifacts and ideas. Between 1990 and 2005, there have been approximately 90 studies that have dealt with analyzing ideas [74]. There are very few studies that deal specifically with assessing creativity by reviewing artifacts [7, 28, 32, 37, 68, 73, 75-80]. To the knowledge of the authors, there are no prior studies that focus specifically on reviewing *ideas for artifacts* (*i.e.* product design ideation). The studies that deal with ideas are not ideas for products and the studies that deal with artifacts involve previously constructed products. In the

case of product design idea generation, designers develop many concepts for products in the form of quick sketches. The review process for this must take from both the studies that deal with reviewing many ideas as well as the studies that deal with reviewing physical products.

2.4.2.1 The Criteria Problem: What constitutes a creative product?

The "criteria problem" is the term given to the issue of obtaining a valid assessment of the level of creativeness of a person [7]. In the studies that have dealt with product ratings, there are various sets of criteria that have been suggested as to what determines a creative product and thus a creative person.

There are two main categories of definitions for creative products: novelty-based and mutli-attribute based [74].

Multi-attribute based definitions of creativity claim that products must be novel as well as possess other quality attributes. The most commonly used quality attribute to describe a creative product is *usefulness* [74]. In engineering design creativity, practicality is an important quality attribute [81]. Mednick defines creative thinking as "...the forming of associative elements into new combinations which either meet specified requirements or are in some way useful. [3]" Other quality attributes include relevance, appropriateness, clarity, workability, feasibility etc. Almost every study in this field chooses different attributes to determine creativity.

Novelty-based definitions claim that products are creative based solely on originality and do not depend on any other quality factors such as appropriateness, usefulness or applicability. Runco and Charles found that "it is not necessary for an original idea to be appropriate to be viewed as creative" [36]. In a more recent study, Christiaans found that "usefulness seemed not to be important in discriminating between designs with high and low creativity ratings [37]".

Amabile takes a stance that is divergent from these attribute-based assessments and claims that creativity cannot be determined objectively using metrics. She suggests a subjective Consensual Definition of Creativity: "A product or response is creative to the extent that appropriate observers independently agree it is creative [68]." In other words, instead of determining if a product is creative by asking reviewers if the product is novel, useful (etc.), simply ask reviewers if the product is creative.

Table 2 presents relevant studies and the dimensions they suggest can be used to determine the creativity of a product (idea or artifact).

Table 2. Summary of Studies Involving the Evaluation of Creative Products or Ideas

Study	Summary	Instrument	Products in Review	Dimensions of Creativity				
Dean et al [74]	Review of 90 Constructs for Idea Evaluation		Ideas (*e.g.* increase tourism in Tucson)	Novelty (originality, paradigm relatedness)	Workability (acceptability, implementability)	Relevance (applicability, effectiveness)	Specificity (implicational explicitness, completeness and clarity)	
Besemer and O'Quin* [76, 77]	Objective Metric for Creative Product Evaluation	Creative Product Semantic Scale (CPSS)	Artifacts (chairs)	Novelty (surprising, original)	Resolution (logical, useful, valuable, understandable)	Elaboration and Synthesis (organic, well-crafted, elegant)		
Shah and Vargas-Hernandez [32, 81]	Evaluation Of Mechanical Engineering Designs		Artifacts and Ideas (mechanical devices)	Novelty (unusual)	Quality (meets specifications)	Variety (explored solution space)	Quantity	
Horn and Salvendy* [78, 79]	Consumer Based Assessment of Product Creativity	Product Creativity Factors	Artifacts (chairs and lamps)	Novelty (frequency, rarity)	Importance (relevance, significance)	Affect (appeal, desire, attraction, delight, stimulation, etc)		
Amabile**[68]	Subjective Assessment Method of Creativity	Consensual Assessment Technique (CAT)	Artifacts (artwork and poetry	Creativity (as determined by appropriate judges)	Creativity Cluster (novel material use, novel idea, effort, detail, etc.)	Technical Cluster (technical goodness, organization, neatness, etc.)	Aesthetic Judgment (liking, aesthetic appeal, would you display it?)	
Christiaans** [37]	Creativity as one metric of design review		Artifacts (cabinets and telephone booths)	Creativity (as determined by appropriate judges)	Technical Quality	Attractiveness	Interest	Goodness of Example

*In these studies, each of the dimensions are made up of many sub-dimensions in the form of bipolar adjective scales **In these studies, the assessment dimensions in addition to "creativity" do not determine creativity, but were used to find correlations with creativity

2.4.2.2 Quantity of Ideas

When dealing with many ideas, as in a brainstorming session, quantity (sometimes termed fluency or productivity) is a common dimension that is reviewed in addition to the dimensions used to measure idea quality. For the most part, quantity is objective and easy to measure. In most studies, a count of non-redundant ideas constitutes the rating for this dimension.

It was said by Linus Pauling , theoretical chemist and biologist, that "the best way to have a good idea is to have a lot of ideas". Research has shown a correlation between total number of ideas and total number of good ideas [64]. Another study found that quantity of ideas was positively

correlated with original ideas and negatively correlated with feasible ideas [66]. Dean *et. al.* found that 18 of the 90 idea review studies used quantity as the sole means of evaluating the ideation session [74]. One could argue that those individuals that are uninhibited when expressing ideas will produce many ideas and their output should be less restrained and thus potentially more creative. It could also be argued that individuals that produce a lot of ideas are better at divergent thinking, which is often related to creative thought process [69, 70]. "The greater number of associations that an individual has to the requisite elements of a problem, the greater the probability of his reaching a creative solution [3]."

Osborn claims that "quantity breeds quality in ideation" and that "early ideas are unlikely to be the best ideas generated during an ideation session" [63]. Creativity tests that ask for the first and only product of an exercise to be judged for its creativity would then be considered a less accurate measure [73]. The relationship between idea quality and idea quantity can be visualized with a curve called the ideation function [82]. The Bounded Ideation Theory (BIT) describes the ideation function as a positive s-curve where the majority of good ideas will come somewhere in the middle of an ideation session [82]. Briggs and Reinig claim that there are three factors that influence the ideation function: understanding the task, cognitive inertia, and exhaustion [83]. When one begins ideation, they are becoming familiar with the task and thus the ideation function has a small ramp-up where there are few good ideas relative to the number of ideas. The ideation function then becomes curvilinear with a positive but decreasing slope and tapers off as the members reach points of cognitive inertia and exhaustion. Cognitive inertia is the term that describes an individual's inability to readily switch to a new and more productive line of thinking [82]. Brainstorming sessions are typically limited in time to avoid cognitive inertia as well as physical and mental exhaustion.

Some studies evaluate only one product/idea per person or group, which makes it easier for a reviewer to evaluate those few ideas on several dimensions. When there are many ideas in need of review, the number of dimensions for which each idea is evaluated must be limited to prevent reviewer fatigue. Amabile found that levels of inter-judge agreement depends on the magnitude of effort required by the judges [68]. In other words, the more products the judges are asked to review, the less the judges agree on ratings.

2.4.2.3 The Judging of Product Creativity

The only requirement of judges is some basic level of familiarity with the subject matter [68]. Expert judges are only required when the products in review are of a highly specialized nature. "The more 'cutting edge' a product is in a specific domain, the more likely an expert judge will be required" [73]. In the case of common household item (such as toaster, umbrella or toothbrush) brainstorming, almost anyone could be considered an appropriate judge of creativity. A few studies have found, based on inter-rater agreement, that there is little difference between experts and non-experts in rating design creativity [37, 68]. Although professionals or experts are capable of producing creative products, they may be unreliable in judging them as they rely on higher, esoteric, or idiosyncratic standards [73]. "For some domains — judging cartoon captions, for example — any "ordinary" individual with an average level of exposure to the written media would be appropriate judges" [68].

In some experiments [22, 66], a panel rates dimensions (such as novelty and practicality) on a scale (e.g. from 1-5). If scores are to be combined for an individual's set of ideas, they will either be added or averaged. If added, a large number of bad ideas are helpful and if averaged, a large number of bad ideas will lower the overall score. It is best to simply count the number of novel, practical, or creative ideas as opposed to average a rating to avoid the influence of a large number of bad ideas [84]. An argument against the count method is that it does not take into account the degree of the creativity of the ideas.

In some cases, the judges agree upon a standard definition for each numerical value of the scale. When dealing with the dimension of creativity, Amabile suggests that the judges use their own definitions of creativity as opposed to attempting to define criteria [68].

Table 3 shows details on judges and rating metrics for a select group of studies that involve the ideas from a brainstorming session.

Table 3. Select Studies Involving Brainstorming Session Evaluation

Test	Paper Topic	Brainstorm Theme	Measured Qualities	# Judges	Scoring	Rating Scale
1998 Butler and Kline [47]	Compare ideation techniques	dealing with drug dealing roommate	quantity quality creativity	self	-	-
2003 Shah, Vargas [81]	measuring idea effectiveness	robot designs	fluency (quantity) novelty variety quality (feasibility)	peer and self review	equations with weights	3 Point
2006 Rietzschel et al [66]	nominal vs. interactive brainstorming	ways to improve education in psychology	originality feasibility productivity	2	averages	5 Point
2006 Reinig and Briggs [84]	compares different scoring methods	what can be done to resolve the problems of the school of business?	idea "quality" productivity	2 experts, 15 year exp.	count of good ideas (4s)	4 Point

2.5 Humor

"Humor is mankind's greatest blessing."
~Mark Twain

According to the Oxford English Dictionary [19] humor is:

a. That quality of action, speech, or writing, which excites amusement; oddity, jocularity, facetiousness, comicality, fun.

b. The faculty of perceiving what is ludicrous or amusing, or of expressing it in speech, writing, or other composition; jocose imagination or treatment of a subject.

Originally the word humor comes from the Latin word for fluid, *humorem*, as in bodily fluids. In time, bodily fluids began to be associated with moods and thus humor then referred to temperament, as in "good-humored". In the sixteenth century, humor was associated with mainly unbalanced temperament and was used to refer to odd, eccentric persons. These odd people became known as humorists as they made others laugh and not until the mid-to late 19th century the term humorist was used for those that produce the product that is intended to amuse others, known as humor. [20]

Humor is now a very broad term that encompasses anything related to funniness and even amusement.

2.5.1 Theories of Humor

"Humor can be dissected, as a frog can, but the thing dies in the process and the innards are discouraging to any but the pure scientific mind [85]."
~E.B. White

There are three main families of theories that attempt to explain the essence of humor: Cognitive or Incongruity Theories, Social or Superiority Theories and Psychoanalytical or Release Theories [86, 87].

Release, psychoanalytical or relief theories revolve around humor being a release of emotional energy or catharsis. We laugh when we are overwhelmed, nervous or even frightened. Freud says that humor results from the frustration of some expectation that should happen but does not [88]. It is believed that humor began as a means of communication in pre-lingual times, as it signaled the good news that a threat was avoided and the group could relax in safety [23]. This form of humor is still exploited today in television shows and movies where the writers use uncomfortable situations to evoke laughter (e.g. Something About Mary, Meet the Parents, Curb Your Enthusiasm, 30 Rock, The Office, The Ali G Show, Borat, and American Pie).

Social or superiority theories were the first attempts to explain humor and they emphasize the negative, aggressive element. Superiority theory views humor as a means to triumph over or make judgment of other people to highlight one's own superiority [87]. In short, we laugh at others misfortunes and when someone slips on a banana peel. The Three Stooges exploited this form of humor, as they intentionally and repetitively became the "butt of the joke" in their shorts. Bergson, perhaps the most known contributor to the superiority theory, believes that humor, in the form of humiliation, is a means in which society corrects deviant behavior [86]. Philosophers like Hobbes and Plato viewed humor as a vice and a display of lack of wisdom [87]. Until the 1860s, it was considered impolite to laugh in public in the United States as laughter was even then viewed as aggressive antipathy [20]. After that time, there was cultural shift from the superiority theory to the more contemporary cognitive/incongruity theory. Traces of superiority humor remain in jokes related to ethnic groups, gender, and low intelligence. However, the original prototypical superiority laughter tends to arise only in the parodies of evil villains. *Muahahahhaa.*

2.5.2 The Incongruity Theory of Humor

The most commonly referenced and most applicable to this study is the incongruity theory of humor. The incongruity theories are cognitively based and attempt to explain humor with less attention to the emotion or social aspects. The basis of the incongruity theory is that things are humorous when there are "disjointed, ill-suited pairings of ideas or situations or presentations of ideas or situations that are divergent from habitual customs" [23]. In other words, something is funny when two or more things come together that are not expected to do so. Arthur Koestler [5] developed the concept of *bisociation*, which is strongly related to the incongruity theory, but he emphasized how this is applicable to humor as well as artistic creativity and scientific discovery. Bisociation occurs when an idea or situation is simultaneously perceived from two incompatible or disparate frames of reference [20]. In other words, taking two concepts that appear to have nothing in common, but with further inspection, finding non-obvious connections.

Most theorists acknowledge that incongruity is a necessary condition for humor but that incongruity by itself is not sufficient as there are many incongruent occurrences that are not funny [20]. Shultz suggested the incongruity-resolution model that states that the incongruity must be resolved in order for humor to exist [89]. Similarly, Suls formulated an incongruity resolution model in his "Two-Stage Model for the Appreciation of Jokes and Cartoons [24]." In his model, shown as a flow chart in Figure 3, the process is a series of steps that begins with a story set-up in verbal or visual format. This set-up is the first part of a joke and it should not contain any incongruent elements. The listener then makes a prediction of what will follow given the information in the set-up. If the ending follows as predicted, the listener will not be surprised and will probably not laugh (although one study found that people find jokes funnier if they expect the punch line [90]). If the ending does not follow as predicted, the listener will then search for a cognitive rule that will make sense of the incongruent information. If a rule is found then the joke should be taken as humorous. If no rule is found then the listener will most likely be confused by the incongruity (however one study found that young children only require incongruity for humor and that the second stage of resolution is only crucial for adults [91]).
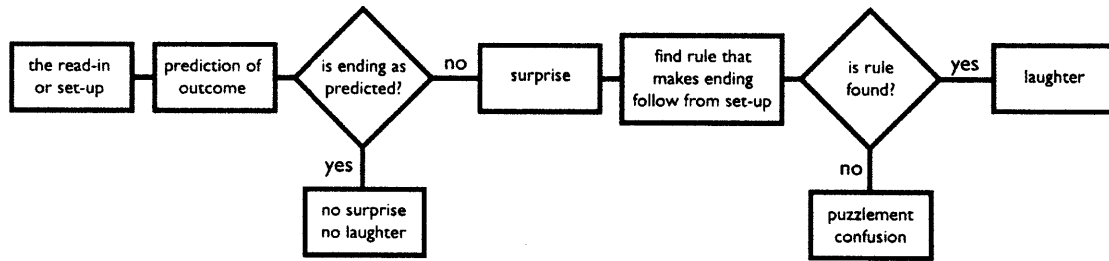
Figure 3. Adaptation of the Two-Stage Model for Joke Appreciatoin by Suls [24]

The rule that needs to be found is the "link," according to Koester, being "the focal concept, word or situation, which is bisociated with both mental planes" [5]. From a linguistics perspective, the punch line or the incongruent ending is termed the "disjunctor" and the rule or link that makes sense of the incongruity is called the "connector" [86]. In a pun, the word (or phrase) that has a double meaning is the "connector". The disjunctor flips the connector's meaning; it is the most important piece of information in the story. The disjunctor is always the final part of the joke and any information after it is superfluous. "There is a continuous series stretching from the pun through the play of words to the play of ideas [5]."

2.5.3 Nonsense Humor

Dogs are forever in the push-up position.
~Mitch Hedberg

Nonsense humor seems to go in opposition to the incongruity-resolution model of humor. Viewing nonsense humor from the incongruity perspective, no rule or connector is present to make sense of the incongruity and the reader is left with a sense of confusion. This is termed a *non-sequitur*. According to Suls, this would typically not result in laughter (in adults). However, if the reader views the lack of resolution as a resolution, laughter should ensue.

Max Eastman, American writer, summarized by Koestler, claimed that nonsense humor is only effective if it pretends to make sense [5]. In essence, this means that the creator must intentionally leave out the resolution with full understanding. Derks found that schizophrenics were not funnier than college students in a test designed to see if individuals with "unusual" thoughts were funny for their incongruent production [92]. The line between confusing-disappointing and confusing-

amusing is very slim and depends mostly on the reader of the joke. Typically individuals who are high in sensation seeking prefer nonsense humor to incongruity resolution humor [93].

2.5.4 Improvisational Comedy

Everyone can act. Everyone can improvise.
~Viola Spolin

Improvisation is performance in which the performers are not following a script or score, but are spontaneously creating their materials as it is performed. The audience observes the creative process in action. In a sense, the creative process is the creative product. [94].

Improvisational theatre began with the Commedia Dell'Arte in Italy in the fifteenth century where there were no definite scripts and performers were free to improvise around familiar situations [95]. This art did not become popular in America until the 1950s when Viola Spolin, a drama teacher from Chicago, developed a series of games and exercises to introduce children to the concept of theatre [94]. Her son, Paul Sils, used these games to begin the first improvisational comedy group, the Compass Players, at the University of Chicago in 1955, which evolved into Second City [94]. Many of Violas games and exercises are published in her book *Improvisation for Theatre* [96].

Generally speaking, there are two types of improvisational theatre: short form and long form. Short form improvisation is short scenes usually constructed from a predetermined game, structure, or idea typically beginning with an audience or MC suggestion [94]. The short form games focus more on developing specific abilities as opposed to the long form improvisation, which focuses on developing plot and characters. Long form is also less focused on comedy than short form [94].

The rules of improvisational theatre have much in common with the rules brainstorming. Just like brainstorming, many short form improvisational games are designed to promote prolific non-obvious association. There is a common rule of improvisation called "yes and…" which essentially sums up the two main rules of brainstorming: defer judgment and build off each other's ideas [28]. Improvisational theatre is a social group activity like brainstorming and the success depends on the participants feeling comfortable sharing ideas and building on the ideas of others. In both improvisational theatre and product design brainstorming, the members of the group share the

common goal of putting out the best product. In improvisational theatre, the goal is not to make yourself look good, rather to make everyone on the stage look good. This means taking what another player suggests and supporting it by adding to it, not by dismissing it and suggesting a new direction. Gerber suggests ways in which several improvisational short form games reinforce the rules of brainstorming and how improvisational games can be used as a brainstorming warm up similar to a warm up in a full body sport [17]. Her work focuses mostly on the benefit of improvisation on improving group interaction.

2.6 Measuring Humor and Wit

Measuring humor is as difficult to measure as creativity. For the most part, this is a result of the lack of agreement on the definition and the scope and the subjective nature of the subject matter.

The goal of many humor tests is to determine if an individual has a "sense of humor" or can comprehend humor. To comprehend a joke and to produce a joke involves different processes. From a linguistics point of view, making sense of a joke is termed disambiguation and it results from making a choice between various meanings (semes) of lexical items (lexemes) that form it [86]. Humor comprehension is a common ability, but wit or humor production is a much less common ability that is regarded as a talent [97, 98]. It is also noted that appreciation of humor is a stable personality trait associated with maturity, whereas wit or humor production can be affected by situational stressors [99].

If dealing with the products of humor (jokes, captions, stories, etc.) or wit. One could adapt Amabile's subjective assessment method; something is humorous to the extent that appropriate observers independently agree it is humorous [68].

2.6.1 "Sense of Humor" Tests

"It is the ability to take a joke, not make one, that proves you have a sense of humor."
~Max Eastman

There is no agreement on what the expression "sense of humor" actually means [100], and so the many tests of "sense of humor" cover many different aspects.

Humor can be used both as a super and sub-ordinate concept. In its subordinate form it can be viewed at the same level as wit, satire, and irony, and can be defined as a "world view allowing one to perceive and enjoy incongruities stemming from imperfections of life and fellow human in a

benevolent way" [100]. In its super-ordinate form it is an umbrella term for all phenomena related to the comic, wit, laughter, and humor (the subordinate form) [100]. "Sense of humor" can be a habitual behavior pattern, an ability, a temperament trait, an aesthetic response, an attitude, a world view or a coping and defense strategy [20]. These "sense of humor" tests that measure the super-ordinate form can cover many diverse areas including: ability to comprehend jokes, ability to express humor, ability to appreciate humor, desire to seek out humor, memory of jokes, and tendency to use humor as a coping mechanism [101].

Ruch [102] classifies 65 humor tests into several categories: informal surveys, joke and cartoon tests, questionnaires and self report, peer report, children humor tests, humor scales in general instruments, and some unclassified. The following are short summaries of some of the more popular tests listed in order of date of publication:

> Sense of Humor Questionnaire (SHQ, SHQ-6) [103] measures a general sense of humor by asking participants to rate or select a level of agreement with different statements. It measures three cognitive and three social items.

> Coping Humor Scale (CHS) [104] measures coping humor by rating self-descriptive statements on a Likert scale.

> The Situational Humor Response Questionnaire (SHRQ) [105] measures the tendency to smile and laugh in a variety of different situations

> 3 Witz-Dimensionen (3WD) [106] measures humor appreciation by asking participants to rate the funniness of a series of jokes and cartoons.

> Multidimensional Sense of Humor Scale (MSHS) [107] measures general sense of humor including coping, appreciation and creativity by presenting a series of statements rated on a Likert scale.

> Humorous Behavior Q-sort Deck (HBQD) [108 1996] measures style of humor by asking about behavior tendencies. Individuals sort cards that each identify a characteristic of a humorous behavior.

Humor Styles Questionnaire (HSQ) [109] measures the ways in which people use humor in everyday life by having participants rate statements on a Likert scale.

The majority of the "sense of humor" tests take the form of questionnaires that touch on humor appreciation, humor comprehension, humor reasoning, humor motivation, etc. These tests have very little in common with wit or humor production [102].

2.6.2 Humor Production (Wit) Tests

Humor production is only one element of the umbrella term "sense of humor." There are very few humor tests that focus on production, and it is the least studied area of humor [110]. Out of the 65 humor tests listed by Ruch [102], the cartoon caption test or cartoon punch line production test (CPPT) was the only one that specifically tested humor production ability.

The following list describes the few tests that were designed to measure wit or humor production ability:

Cartoon Caption or Cartoon Punch line Production Test (CPPT) [10] - involves humorously annotating a set of caption-less cartoons or images.

Wit Selection Measure [14] – involves generating funny statements such as create a funny campaign slogan.

Wittiness Questionnaire and Humorous Monologue [110] – involves three parts: a questionnaire on self rated and peer rated wittiness, a traditional cartoon caption test, and a three minute improvised monologue with props

Geographic Pun Generation [25] – involves coming up with puns related to geography. Details of the test are unclear.

2.6.3 Cartoon Caption Test

"Cartoons are like gossamer, and one doesn't dissect gossamer."
~Jerry Seinfeld

The first use of the cartoon caption test was as a creativity test and not specifically a test of humor production. In this study they asked groups of three male undergraduates to write "as many clever or somewhat amusing captions as possible for this Saturday Evening Post cartoon with a time limit

of five minutes [111]". They scored the ideational fluency (or number of captions) as well as the humor of the captions on a 5-point scale. In 1970, the test was used as a measure of humor production ability where participants were asked to annotate a set of caption-less cartoon images with a humorous caption [10]. The cartoon caption test is the only repeatedly used test for humor production ability. Table 4 summarizes the details of the prior studies that use the cartoon caption test.

Table 4. Summary of Studies Involving a Cartoon Caption Test

Test	# Subjects	# Cartoons / Captions	Time Limit	Total # Captions	# Judges	Rating Scale
1962 Ziller, Behringer, Goodchilds [111]	192 undergraduate males, in groups of 3	1 cartoon / as many as possible	5 minutes	~573	?	5 point
1970 Treadwell [10]	83 science/engineering undergrads	18 reduced to 11, one per	2.5 min / cartoon	913	2	5 point
1973 Babad [11]	77 female undergrads	15, one per	1 min / cartoon	< 1155	13	2 points
1976 Brodzinsky and Rubien [13]	84 undergrads	12, one per	5 min / cartoon	1008	6	6 points
1980 Turner [110]	87 undergrads	12, one per	2.4 min / cartoon	< 1044	2	4 points
1988 Derks and Hervas [112]	38 undergrads	3, 2 or 10 per	NA	684	10	11 points
1993 Feingold [21]	51 and 47 Central Park visitors	8, one per	NA	784	2	5 points
1996 Kohler and Ruch [101]	110 German adults	15, as many per	30 minutes	1650	12	9 points

These prior studies tested between 38 and 110 individual participants. The first six tests utilize only undergraduate student participants. In the majority of these studies, the participants are asked to come up with just one caption for anywhere between 8 and 18 cartoons. One study, by Derks and Hervas, specifically asks the participants to produce either 2 or 10 captions [112].

The studies that mention a time limit allotted between 1 and 5 minutes per cartoon. The total number of captions produced in each study ranged between 573 and 1650. In each study, 2-13 judges rated each of the captions. Rating scales vary between 2 and 11 points. Derks suggests a rating scale between 10 and 20 points as "more would be superfluous; less would not accommodate humor's dynamic range [112]." In reviewing captions, Derks found that sexual and innocent humor were judged in the same qualitative way, but males increased their judgment of funniness by a fixed amount if sexual humor was present [113]

One of the most recent uses of this test involved sample cartoons from the New Yorker Magazine [21]. The New Yorker, published since 1925, created the Cartoon Caption Contest (online at http://www.newyorker.com/humor/caption) in 2005. In the contest, the magazine publishes a captionless cartoon on the last page of each issue for readers to submit and vote on captions. The contest was made into a board game in 2006. An example of a recent New Yorker caption-less cartoon is shown in Figure 4.

Figure 4. Example of a Cartoon from the New Yorker Cartoon Caption Contest

2.6.3.1 Quantity of captions

Most of the prior studies involving caption creation ask the participants to create only one caption per cartoon. A study by Derks and Hervas [112], found that caption funniness improved with output order and that subjects produced better captions when they produced more captions. Derks and Hervas relate this study to creativity by concluding that "early ideas are not as creative as later ones" and " many ideas will produce more good ideas than will few [112]." Derks also found that subjects are most likely to produce the best caption after the worst caption 64.5% of the time [114]. In these studies, captions were rated by a group of 10 peers.

2.6.3.2 Advice from Expert Cartoon Caption Writers

Since the New Yorker released the Cartoon Caption Competition in 2005, there have been a few articles published by winners promoting their tricks and recipes for creating a winning caption.

Patrick House, winner of the New Yorker Cartoon Caption Competition explains that one must design the caption for the intended audience, being "isolated and introspective" readers that would not necessarily laugh out loud. He suggests creating "Theory of Mind" captions instead of puns or

visual gags. "Theory of Mind" refers to higher order jokes that require the reader to "project intents or beliefs into the minds of the cartoon's characters." At the time of this publication 136/145 (94%) of the winning captions would be considered "Theory of Mind." House's final suggestion is to use "common, simple and monosyllabic words" and to "steer clear of proper nouns that could potentially alienate." To relate to the audience, one should use language that everyone understands and can relate to. Proper nouns bring subjectivity and are not timeless. At the time of this publication, excluding first names, only nine proper nouns have appeared in winning captions: Batmobile, Comanche, Roswell, Hell, Surrealism, Tylenol, Bud Light, Frankenstein, Kansas Board of Education. To sum up House's technique: Use language that most people understand, use higher level "Theory of Mind" humor, and don't try to be too funny. [115]

Larry Wood [116], a three-time winner of the New Yorker Caption Competition, has a different strategy. Wood suggests trying to be as funny as possible, even though he says that all of his winning captions were more clever than funny. His two most important tips are to be brief and to incorporate everything that is going on in the cartoon. One theory of aesthetics is that "more is less," where simple is more aesthetically pleasing than complicated. This theory could carry over to humor, where communicating the same joke in less words could be viewed as more funny. As a side note, Wood mentioned that the captions that he submitted that were selected as finalists were not as strong as some of his other captions [116].

Scott McCloud is a comic writer and author of the trilogy of non-fiction books on comics design: Understanding Comics [117], Reinventing Comics [118] and Making Comics [119]. In Making Comics, McCloud describes seven ways of combining pictures and words: Word Specific, Picture Specific, Duo-Specific, Intersecting, Interdependent, Parallel, and Montage. Word Specific refers to instances when the words provide all the meaning and the pictures illustrate aspects of the scene. Oppositely, Picture Specific refers to instances when the pictures provide all the meaning and words only accentuate aspects of the scene. Duo-Specific refers to instances when the words and pictures provide the same information in redundancy. Intersecting refers to instances when the words and pictures provide some similar information, but each provides additional information to the scene. Interdependent refers to instances when the words and pictures combine to create a concept that neither would convey by itself. Parallel refers to instances when the words and pictures follow seemingly different paths that do not intersect. Montage refers to instances when

the word and pictures are combined pictorially in the form of word art, sound effects, graphics, logos etc. [119]

Picture Specific and Duo-Specific responses on a caption test would mean that the participant did not add any new information by creating the caption. Therefore, if no new information is added, it is hard to create incongruity and thus humor. Parallel responses on a caption test would be considered *non sequitur* but might be considered humorous. Aside from the use of onomatopoeia, punctuation and capital letters, Montage responses are not possible in the caption test as the captions are typically typed before review. McCloud seems to favor interdependent combinations as they "keep readers' minds fully engaged because they require them to assemble meanings out of such different parts. Such effects can be stimulating, gratifying and a kind of experience rarely found outside of comics [119]."

2.7 The Cognitive Science of Humor and Creativity

There have been several different studies that address the neurological basis of humor [120], [121] and creativity [22] [122], but none specifically comparing these processes or specifically addressing humor *production*. There are four areas of brain that are discussed often when dealing with the cognitive aspects of humor and creativity: the right hemisphere, the frontal lobe, the temporal lobe, and the limbic system (specifically dopamine production) [122].

2.7.1 Dopamine

Dopamine is closely related to creative thought as it plays a large role in the drive to be creative and make associations [122]. People with low amounts of dopamine are typically depressed and can develop Parkinson's disease. People with too much dopamine are psychotic and take associative thought to a higher level making ridiculous associations between things and accepting them to be true. Things that stimulate dopamine production have been found to stimulate innovative thinking and wittiness, such as marijuana [123] [124] and positive affect from candy and funny movies [58]. Marijuana has shown to promote free association and increase the amount of novel and uncommon associations [123].

Humor, in response to funny cartoons as well as comprehension of puns and jokes are in close relation with creativity, the right hemisphere and release of dopamine [22].

2.7.2 The Right Hemisphere

Generally speaking, the right hemisphere tends to be highly active in both humor and creativity studies. It governs primary process cognition, which facilitates discovery of new combinations of mental elements. Participants that were more creative in an alternative uses test had higher activity in the right hemisphere than the left [22]. The right hemisphere governs perception, music production, visual art creation, and our language lexicon [4]. Hypnosis [124] and marijuana [125] have both been proven to increase right hemisphere activation as well as creativity. Even looking leftward instead of rightward was shown to increase creativity, as the brain processes the information first with the right hemisphere [126, 127].

The right hemisphere plays a crucial role in both the processing of humor and the resolution of humor [128]. A study by Shammi and Stuss found that patients with damage to their right frontal lobe were not able to appreciate humor [129]. Brownell *et al.* found that right hemisphere damaged patients were able to find the incongruity in a joke and be surprised, but were not able to make the resolution to comprehend the punch lines [130, 131].

2.7.3 The Frontal Lobe

The frontal lobe was found to be associated with the generation of original ideas as well as divergent thinking [22]. Creative storytelling and making distant associations takes places in the prefrontal areas of the right hemisphere as well as the anterior cingulate cortex (ACC) [132]. Another study found that the left inferior frontal lobe is active when making any semantic associations, but the right frontal lobe was active in making unusual or distant semantic associations [133]. Magnetic stimulation over frontal lobes was found to increase creativity in certain tasks [134].

2.7.4 Right Posterior Superior Temporal Sulcus (PSTS)

The closest fMRI study to understanding both humor production and innovation investigates what happens in the brain when novel connections are made between unrelated words [135]. This process is termed "verbal creativity" and it takes place in the right posterior superior temporal sulcus (PSTS) shown in Figure 5 [135]. "Verbal creativity" is the basis of creative thought according to some [3-5] as well as the basis of wit and humor comprehension according to others

[24, 87, 136]. In separate study, this area was found be involved in "conceptual matching, irrespective of whether the stimuli are audiovisual, auditory–auditory or visual–visual" [137].

The PSTS is the right hemisphere's version of Wernicke's area [135]. Wernicke's area, which typically resides in the left hemisphere, needs to function properly for successful linguistic creativity [122]. Damage to this area is termed Wernicke's aphasia and results in meaningless speech with normal sounding rhythm and syntax. Adjacent to the PSTS is the right anterior superior temporal sulcus (ASTS), which was found to be highly active in the "Aha!" experience, which occurs after making distant connections [138].
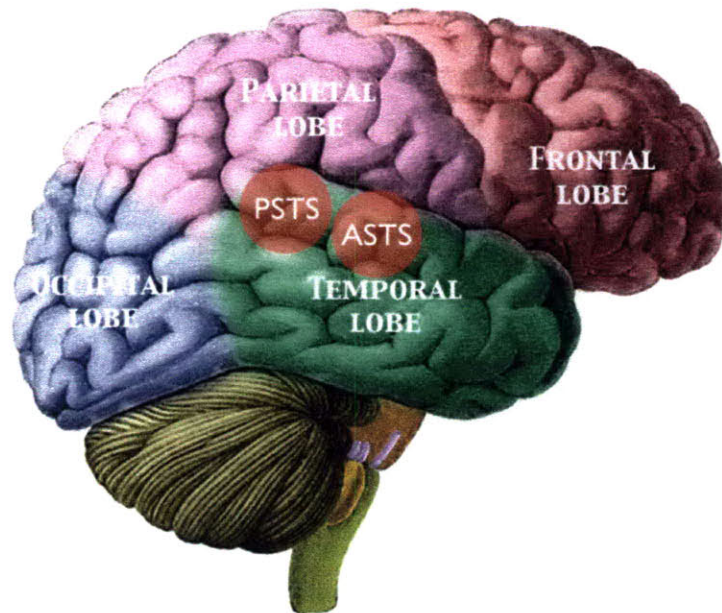
Figure 5. Brain viewed from the right side showing the 4 major cerebral lobes and the locations of the Posterior Superior Temporal Sulcus (PSTS) and the Anterior Superior Temporal Sulcus (ASTS). The original illustration is from Manuel de L'anatomiste, by Charles Morel and Mathias Duval, published in 1883 digitally enhanced by Scott Camazine.

2.7.5 Improvisation

The few studies that have addressed improvisation from a cognitive science perspective have focused on music and dance. It was found that jazz musicians, while improvising, have brain activity comparable to REM sleep and dreaming, in which they deactivate part of their brain that is responsible for goal-directed actions and rational association [139] [140]. Another paper explored the difference between professional dancers and novice dancers while imagining improvised dance and found that professional dancers exhibited more right-hemisphere alpha synchronization than

46

novice dancers [141]. In this study the professional dancers also exhibited higher alpha synchronization in posterior parietal brain regions than the novice dancers during an alternative uses test [141].

2.8 Gaps in Research

2.8.1 Humor Production and Product Ideation Correlation

The studies that have involved humor and creativity correlations are few and dated. Out of these studies, even fewer specifically focused on humor production, which the author believes to be the strongest connection to the realm of creativity. The creativity tests that have been used in these prior studies, such as the RAT and the TTCT, are tests of specific thinking abilities such as divergent and convergent thinking. None of these prior studies examine applied creativity or what would be called "product rating tests." Product ideation or brainstorming is a process that is used in industry, and therefore correlations with a test of this process would have more meaning than a standardized creativity test. It is hypothesized that individuals that are fluent humor producers should also be fluent in innovative product ideation.

Prior studies have not explored the correlation of humor production to attributes of creative products such as novelty and usefulness. It is hypothesized that individuals that are fluent in humor production produce high amounts of novel product concepts but few useful concepts.

2.8.2 Fluency

Few studies have addressed humor fluency as a measure and component of wit. It is hypothesized that individuals that produce more captions should also produce more funny captions in a cartoon caption test. Almost all prior cartoon caption studies ask the participants to produce one caption per cartoon. We believe that this does not capture an individual's true humor production ability. In this study, we will ask participants to write as many captions as they can per cartoon. As no prior studies compare ideation to humor production, we can explore how fluency in humor production compares to fluency in idea generation.

A few studies try to prove or disprove Osborn's theory that more ideas will result in more good ideas. We can use this study to explore how quantity correlates with quality in idea production specific to product design ideation.

2.8.3 Evaluating Product Ideas

There are many tests of creativity and relatively few tests of creative production. Ideation or idea generation is rarely used as a measure of creativity. In the studies that do evaluate an idea generation session, none involve product concepts. Typically, tests that evaluate ideas involve prompts related to situational problems (i.e. increasing tourism in Tucson, improving the psychology department at a college, or dealing with a drug dealing roommate). These situational problems are not applicable to creativity in the realm of design, specifically product design. These questions are also highly structured where there can be "correct" responses and usefulness has a large impact on the goodness of the responses. This study will set the groundwork for a test of product design creativity.

There are over 80 studies that evaluate the creativity of a product or idea and they rarely use the same instrument of metrics for rating or determining creativity. A large group of these instruments are specific to ideas that are not product ideas. Oppositely, few are specific to existing products and not ideas for products. These few studies review a small selection of products and not hundreds. A hybrid of these instruments is needed to evaluate product ideas from an ideation session. This study will create an instrument that can be used to quickly evaluate a large quantity of product ideas. We can also see which metrics correlate with subjective ratings of creativity and develop a tool to better classify product ideas.

2.8.4 Review Process for Ideas and Jokes

The prior studies that have dealt with the rating of humor and ideas have a small select number of judges (at minimum 2). Some studies use "expert" judges to rate ideas or humor. It is believed that 2 judges are not appropriate as individuals have different tastes in humor and products. For this study we experiment with an online review process so we can quickly average subjective assessments from many reviewers.

It is believed that clarity and drawing ability can influence perceived creativity. Similarly, it is believed that the wording of a joke plays a large role in the perceived humor. Prior studies have not explored correlations of presentation (wording of humor and sketching of ideas) with scores for humor and creativity.

2.8.5 Learning from Industry Professionals

The studies that have dealt with humor production or idea generation have not compared different demographics including those that would be considered experts in the areas of humor or idea generation. In this work we explore differences in humor production and product ideation between professional designers, students, and improvisational comedians. We can also look at the effects of gender on humor and idea generation.

We believe that creativity and wittiness can be learned. A few researches have suggested that designers can benefit from improvisation training as there are many similarities between that and brainstorming, but none have collected empirical data. These studies have also focused on the group interaction aspect of improvisation training and not the association making.

3. Experiment

The basis of this study involves a series of correlations between individual performances on a cartoon caption test and performance in a nominal product brainstorming. A second study builds off of this foundation, which involves testing the effect of improvisational training on idea generation.

3.1 Pretests

Several pretests were performed, including: a nominal product brainstorming; a cartoon caption test; a joke punch line test; a Remote Associates Test; and a practice improvisation workshop.

Both the nominal product brainstorming and cartoon caption pretest were conducted with four participants (two male and two female). The ages of these pretest participants ranged from 19 to 28. These participants were all MIT students with minimal to high-level experience with product design. These four participants were briefed with the goal of the research and were given the cartoon caption test and nominal product brainstorm test on different days. There were several goals of this pretest: 1) determine which of the many cartoons and product themes would be most appropriate for the actual testing, 2) determine how many cartoons and products to use in each test, 3) determine a proper length of time for each test, 4) see if there are any correlations in this small sample size, and finally 5) practice a means of administering the test and collecting and reviewing data.

Several measures were taken to ensure a comfortable testing environment. Testing was performed in familiar environments with ample table space and mostly comfortable seating. The administrator was present during the testing to observe and answer questions, but did not sit with or near the participants. These tests were also administered in a group setting and so the participants were told not to talk or communicate with the other participants.

In reviewing the pretest results, two different sets of individual judges were used. Four male and two female judges (age range between 23 and 45) reviewed the captions, and three male product design instructors (age range between 25-45) reviewed the product ideas.

The remote associates pre-test and the second cartoon caption pretest were conducted at a later time than the other pretests.

3.1.1 First Cartoon Caption Pretest

In determining which cartoons to use for the caption test, we reviewed prior studies that used this test [10, 11, 13, 21, 112], as described in Table 4. None of these studies provided pictorial examples of the cartoons that were used, however, the most recent study made reference to the New Yorker Cartoon Caption Contest. In 2006, the New Yorker released a board game that consisted of a deck of cards containing caption-less cartoons featured in past issues of the magazine. We decided to choose our cartoons from this collection. In a first pass of the cards, we removed any that were potentially violent, sexual, or offensive in nature. Violent or sexual themes were found to create a bias towards males in humor production abilities [13]. Somewhat comically, the deck of 189 cards was reduced to 12 that showed potential. These cartoons are shown in no particular order in Figure 6 listed from A to L.
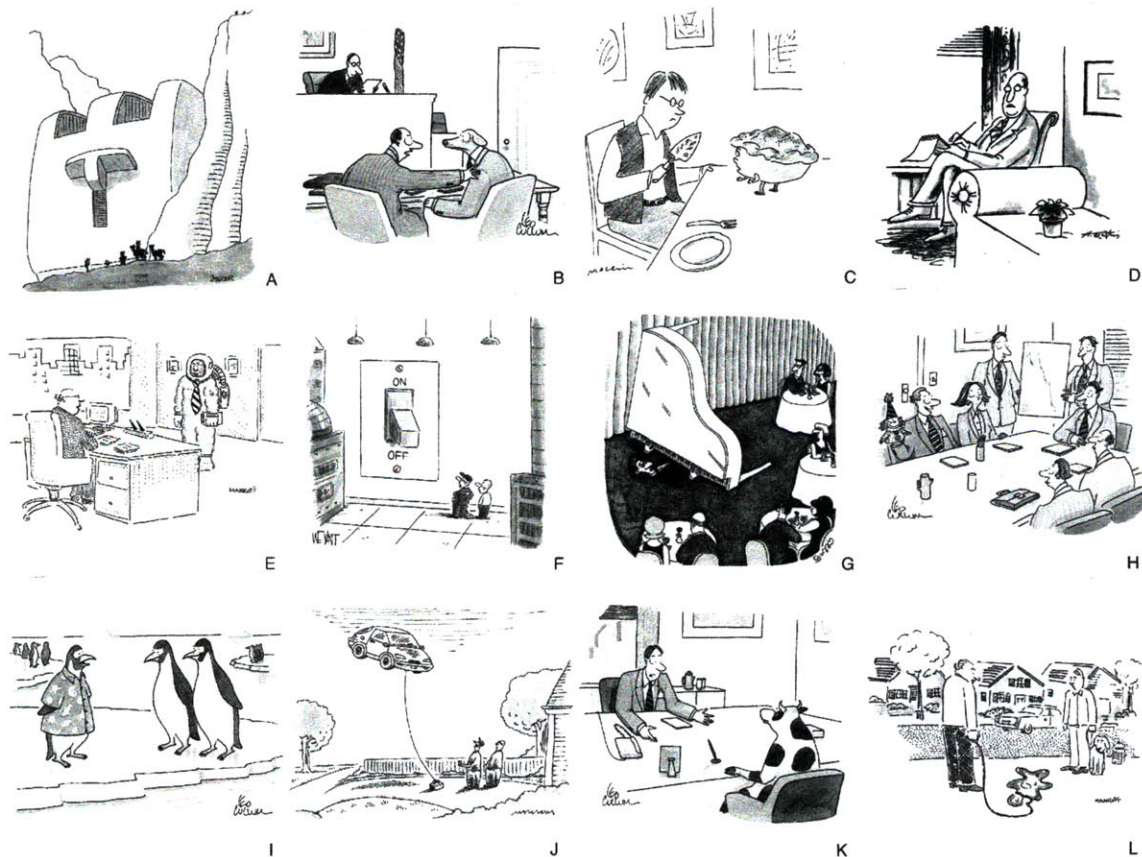
Figure 6. Selection of Cartoons from the New Yorker Cartoon Caption Game

Each cartoon was printed on a card with a letter designation and participants were given the cartoons in a varied order. Each participant was given a ruled pad and a pen to write down captions. Participants were told to write the letter of the cartoon on the page as well as their identification number. Participants were told to start a new page of captions for each new cartoon. Participants were prompted with the following:

> This is a study in humor production ability. You will be shown a caption-less cartoon and we would like you to write down as many humorous captions as you can in the next five minutes. Please write down any captions that come to mind even if they are only mildly amusing, silly or offensive. Please write legibly and start each new caption with a bullet point. After 5 minutes, you will be asked to finish any caption that has been started and we will give you another cartoon.

Warning the participants of how much time they have left may interrupt the creative process. Participants were told to write as many captions as they can even if they feel they are not funny. Derks found that participants that were asked to write a high quantity of captions had funnier captions than those that were instructed to write only high quality captions; those instructed for quantity over quality also had more unfunny captions [114].

Originally, participants were given 3 minutes for each cartoon prompt, but we found that 5 minutes was more appropriate. A previous study [13] also allotted 5 minutes per cartoon. Participants were given 10-minute breaks between sets of four cartoons to prevent fatigue.

As some participants may be more comfortable or familiar with different cartoon themes, a larger set of cartoons would provide better data, however, as the number of cartoons increase, the participants become more fatigued. Participants of the pretest stated that they became restless after three or four cartoons. We decided that three cartoons was an appropriate number for the experimental testing.

For reviewing, all captions were typed and stripped of any identity to the creators. Six reviewers judged the captions (2 female and 4 male). Reviewers were asked to read all the captions for a

given cartoon and give each caption a score of 0,1, or 2.   A score of 0 was defined as not funny, a score of 1 was defined as amusing or mildly funny, and a score of 2 was defined as genuinely funny.

With the scores for each cartoon, the author generated three values. The first was the average number of responses per participant for each cartoon, the second was the average number of funny captions created (*i.e.* average number of captions given a score of 2 per reviewer) for each cartoon, and the third is the percentage of funny captions (i.e. the number of captions that were scored 2 out of the number of captions).

Table 5 presents these scores for each cartoon along with an unofficial cartoon title to make referencing easier.

Table 5. Pretest Scores for Selection of Cartoons

		AVG # of Responses	AVE # of 2-Scores	Percentage of Funny
A	Giant Toaster	4.3	3.0	18
B	Dog in Courtroom	5.8	2.8	12
C	Talking Pie	5.3	2.8	13
D	Plant Psychologist	6.5	3.8	15
E	Spacesuit in Office	4.8	2.7	14
F	Giant Light Switch	6.8	3.5	13
G	Piano Tip	4.8	2.3	12
H	Puppet in Meeting	5.3	2.2	10
I	Hawaii Penguin	6.8	3.0	11
J	Balloon Car	7.5	3.8	13
K	Cow in Office	5.5	2.3	11
L	Amoeba Walking	6.5	3.3	13

In evaluating the captions, we did not average the scores, but rather counted the number of high scores.   This is done to reduce the negative effect of a large number of unfunny captions [84].

The author wanted to choose the cartoons that allowed for the most humorous captions. Having a high number of responses is good, but more importantly is the number of funny responses or the percentage of funny captions.

After eliminating all captions that did not score well in any of the three categories (B, C, G, H, K), a few additional cartoons were eliminated based on types of responses: (L) Amoeba Walking produced a high number of death related responses, (F) Giant Light Switch produced mostly

engineering and science related responses (which could be related to the background of the participants), and several participants found (J) Balloon Car to be confusing.

Of the remaining four cartoons, A (Giant Toaster), D (Plant Psychologist) and E (Spacesuit in Office) had the highest percentage of funny captions and these were the three cartoons that were chosen for the final experimental testing. These cartoons are shown in Figure 7. In retrospect, perhaps the count of high-scoring captions is more important than the percentage of high scoring captions and so cartoon I (Hawaii Penguin) should have been used instead of cartoon E (Spacesuit in Office). This observation was made well after the pre-test. The author believes that both cartoons E and I are acceptable for this study.

(a)　　　　　　　　　　　　(b)　　　　　　　　　　　　(c)
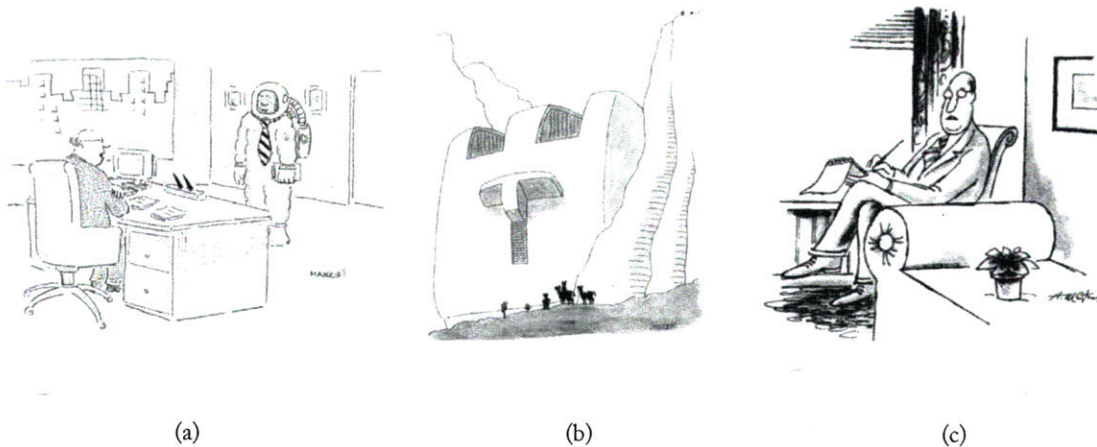
Figure 7. Three Test Cartoons

The following are the original New Yorker captions for these three selected cartoons.

> Figure 7a (Spacesuit in Office) was drawn by Robert Mankoff and published in the New Yorker on January 17, 2000 with the original caption: "You know, Burkhart, if you're so damn afraid of the flu maybe you should just stay home."

> Figure 7b (Giant Toaster) was drawn by Jack Ziegler and published in the New Yorker on April 18, 2005 with the original caption: "At long last, Wyatt, our dream has come true and we are within reach of the legendary toast fields of the Sierra Madre."

> Figure 7c (Plant Psychologist) was drawn by Henry Martin and published in the New Yorker on May 11, 1992 with the original caption: "Do you think your failure to bloom

could be caused not just by improper location but also by a fear of having your blooms compared with those of other African violets?"

These original captions were included anonymously as part of the review as a "control" to see how the published captions compare to those of the participants.

The authors believe that for a small number of items to rate, a scale of 10 to 20 points could be acceptable as per Derks, but when a reviewer potentially rates hundreds of captions, the reviewer cannot keep a constant mental rubric unless the scale provides qualitative descriptions for each number value. One prior study rated on a binary funny/not funny, which we believe is not descriptive enough. A 1-5 or 0-5 point scale is what the majority of the prior caption tests have employed. We will use a 3-point scale with brief qualitative descriptors for our final experimental study.

3.1.2 Second Cartoon Caption Pretest

A point of interest was whether or not the participants in the caption test should be asked for one funny caption per cartoon or as many funny captions per cartoon as possible. Five of the eight studies in Table 4 involving the cartoon caption test ask the participants to produce just one caption per cartoon [10, 11, 13, 21, 110]. It is believed that a creativity test should not be based on the first and only product [73].

It is believed that quantity breeds quality in idea generation [3, 63, 64] and thus it should also be the case that more captions produces more funny captions. It was theorized that the majority of good ideas do not come at the beginning of an idea generation session and more towards the middle of a session before participants become exhausted [82]. Derks found that caption funniness improved with output order and that subjects produced better captions when they produced more captions [112].

In a second pretest, we performed the cartoon caption test with a group of 21 incoming freshmen girls. Each girl was given the same cartoon (Figure 7a) and asked to come up with as many captions as they can while following the guidelines presented in 3.1.1 First Cartoon Caption Pretest. After completion of the test, they were asked to circle the caption that they felt was the most funny.

The average number of captions produced per participant was 6.1. The average placement of the self-selected funniest caption in order of creation was 3.3. This is a first order estimate that the funniest caption (or at least the caption that the creator feels to be the funniest) will occur in the middle of the testing. Figure 8a shows the placement of the self-selected funniest caption. This data alone is misleading as different participants produced different numbers of captions (ranging from 3 to 10 captions). However, one can see that in only 4/21 (~19%) of the instances, the participants felt that their first caption was the funniest. We can also present the data in the form of an average percentage into the testing. Figure 8b shows the placement frequency of the self-selected funniest captions. The majority of funniest captions occur in the third fifth of caption production.

(a)   (b)
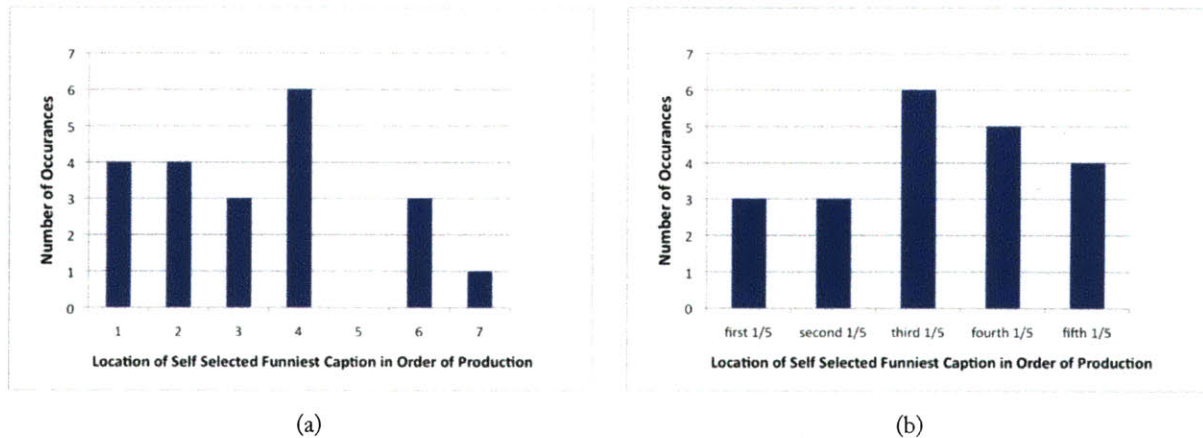
Figure 8. Placement Frequency of Self-Selected Funniest Caption

The data presented in Figure 8b can be compared to the Bounded Ideation Theory (BIT) [82]. The findings show few funny captions in the beginning as the participants are become familiar with the prompt and testing. The majority of funny captions in the middle with a taper towards the end as the participants become exhausted and cognitively fixated on similar concepts. To get a true testing of individuals humor production abilities, we will ask the participants to ideate as many humorous captions as they can as opposed to only one caption per cartoon.

3.1.3 Punch Line Completion Pretest

As there only appears to be one repeatedly used test of humor production, we wanted to create another humor production test. We developed what we call the Punch Line Completion Test. In

this test, participants are told the set-up of a common joke, which has many options for customization, and then they are asked to come up with as many funny punch lines as they can.

The joke set-up we chose was:

"A _____ walks into a bar and the bartender says I can't serve you because..."

The blank space allows for several different set-ups with different subjects walking into a bar. The participants then create punch lines in the form of why the bartender cannot serve the subject of the sentences. We tested several different subject matters including: elephant, magician, computer, penguin, caveman, pirate, pumpkin, lawyer, doctor, dancer, watermelon, and Santa Claus.

This joke in particular was used, as it is a common format for a joke that can be manipulated easily. The fact that it involves a bar brings with it a tone that anything is appropriate. It is also quite short and typically leads to associations involving the subject of the joke and something related to a bar. A version of this test has been used previously to determine humor perception, and not creation, where a "set up" line was provided and the participant chooses the funniest punch line out of four options [129].

In our pretesting, we asked six participants (3 male, 3 female) to produce as many punch lines as they can in three minutes for each of these following subject matters: elephant, magician, computer, penguin, caveman, and pirate. Participants stated that the test was very challenging and that their jokes were not funny. After having two judges read all of the punch lines, they too agreed that none of the jokes were funny. Perhaps this test asks the participants to come up with a very specific joke, which does not allow for much creative freedom. We decided to use only the cartoon caption test as the measure of humor production ability as it has been used reliably to produce humor in prior studies and in our pretests.

3.1.4 Pretests of Nominal Product Brainstorming

The purpose of this study is to determine if there is a correlation between humor production and creativity in the field of product design. Instead of correlating humor production with a standardized creativity test (such as the RAT or the TTCT), we decided to have the participants ideate product concepts for a given product theme. There are several reasons for this decision: 1)

creativity tests are measures of specific thought processes and not necessarily those that are crucial for product ideation, 2) the brainstorming method is used in practice and thus it is much more meaningful than a creativity test, 3) creativity depends on the product not on the task as per Amabile [68], and 4) there are prior studies that attempt to correlate the cartoon caption test with standardized tests of creativity (RAT, TTCT), but none with a more practical test.

We chose to have the participants ideate around a product theme (ill-structured) instead of developing solutions to a problem (structured). There are several reasons why we chose an ill structured or blue-sky idea generation over a structured theme. In structured problems, the solutions may not be related to products at all and it would then be hard to compare innovative designs. Designers and engineers may develop better solutions to structured problems as a result of their training in the scientific method and less because of creative abilities. The structured problem prompt also suggests there is in fact a solution or best solution and we are trying to avoid that type of thinking. We wanted an ill-structured prompt so as there are no correct or best answers [47]. Finally, in our recent experiences, consultants often come to us asking for blue-sky innovation around a product theme such as peristaltic pumps, water guns, or cork. These are real scenarios. In the future, it may be beneficial to look into a structured problem ideation as this may involve a different thought process.

In the nominal brainstorming pretest, participants sat together at a large table without talking or sharing ideas. We decided to use a nominal brainstorming opposed to a traditional interactive brainstorm so we can better evaluate the individual participants. Participants were asked to develop as many innovative concepts they can around a given product them. The concepts could be related to that theme in any way. They were asked to sketch each idea with a title on a new piece of paper in the portrait orientation. We asked that all sketches be made in the portrait orientation to simplify the scanning, processing and reviewing. Participants were asked to draw legibly and use as few words as possible to explain the concept. They were told that drawing ability does not count and that they should sketch all ideas that they feel are innovative in some way

To determine which products should be used in the final testing, we began with several household products that are familiar to the general population. These products should be common enough for the general public to have a good understanding of the current state of the art to be

comfortable enough to innovate around that theme. These products should not be engineering specific to avoid a bias toward engineering innovation. The products should be common enough for any "average" person to be considered an appropriate judge of creativity [142]. We also avoided products that would tend to inspire offensive or sexual themes such as beds and toilets.

We originally chose six product themes: toasters, umbrellas, toothbrushes, trash bins, lamps, and pens. Coincidentally, a similar study evaluating the creativity of products chose three of these same themes (toasters, toothbrushes, and lamps) because of their "commonality, moderate cost, and level of interest for general young professional consumers [79]." This similar study evaluated products that were currently on the market and not conceptual product ideas.

The pretest group of four students was given 15 minutes for each of these six product themes in varied order with a 10-minute break between sets of three. The number of ideas produced per participant and theme is shown in Table 6. The average number of ideas produced per concept per participant was 6.7.

Table 6. Average Number of Ideas per Product Theme

Theme	Average # Ideas / Person
Lamp	4.8
Pen	8.5
Toaster	7.3
Trash Bin	4.3
Toothbrush	8.0
Umbrella	7.5

We found that 15 minutes was quite draining for the participants, especially after six rounds. The participants expressed that most of their ideas did not occur in the last few minutes and that 12 minutes would be acceptable. To keep consistency with the cartoon caption test and to keep the entire test under an hour in length, we decided to use three product themes for the final testing.

In deciding which of these themes to use in the final experimental study, we first eliminated trash bins and lamps as the participants struggled to produce about half the number of ideas as produced for the other four themes. For the remaining four themes, we looked to number of innovative concepts and percentage of innovative concepts.

In reviewing the ideas, we asked three product designers to rate each concept on three categories: the novelty (how original is the idea) from 0-2, the usefulness (how useful is the concept) from 0-2, and the funniness of the idea from 0-1. We kept track of the ideas that were given both a high score on novelty and usefulness and we counted these ideas as innovative. This reviewing method and definition of innovative changed slightly for the final testing, but for the pretest, this method provided a first order estimate of the innovativeness of the concepts. The average scores are shown in Table 7.

Table 7a represents the average number of high scoring concepts per person, and Table 7b presents this as a percentage of the total number of ideas created for each theme.

Table 7. Average Participant Scores in Pretest of Product Brainstorming

	Novelty	Practicality	Innovative	Humor
umbrella	2.9	2.6	1.3	2.9
toaster	2.5	2.7	1.3	2.3
pen	2.6	2.8	0.8	3.0
toothbrush	3.1	2.4	0.9	2.2

(a)

	Novelty	Practicality	Innovative	Humor
umbrella	39%	34%	17%	39%
toaster	34%	37%	18%	32%
pen	31%	32%	9%	35%
toothbrush	38%	30%	12%	27%

(b)

One can see that the pen theme had the least number and percentage of what we designated as highly innovative concepts. It also had a low novelty score and the lowest percent of novel ideas. We decided to discard the pen theme and use umbrellas, toasters and toothbrushes as our three product themes for the final experimental study. It is also interesting to note that a toaster appears in the cartoon caption test as well.

We found that some concepts were often hard to understand and that some used more text than images. In our actual testing, we emphasized legibility, minimal word usage and large sketches. An identification number specific to each participant was placed on the backs of each concept drawing and each drawing was scanned.

3.1.5 Humor and Creativity Pretest Correlations

The pretest was performed primarily to decide on proper testing format and selection of testing material. Although there are only four participants and small set of reviewers, we can also look for any interesting trends in the data.

(a)   (b)
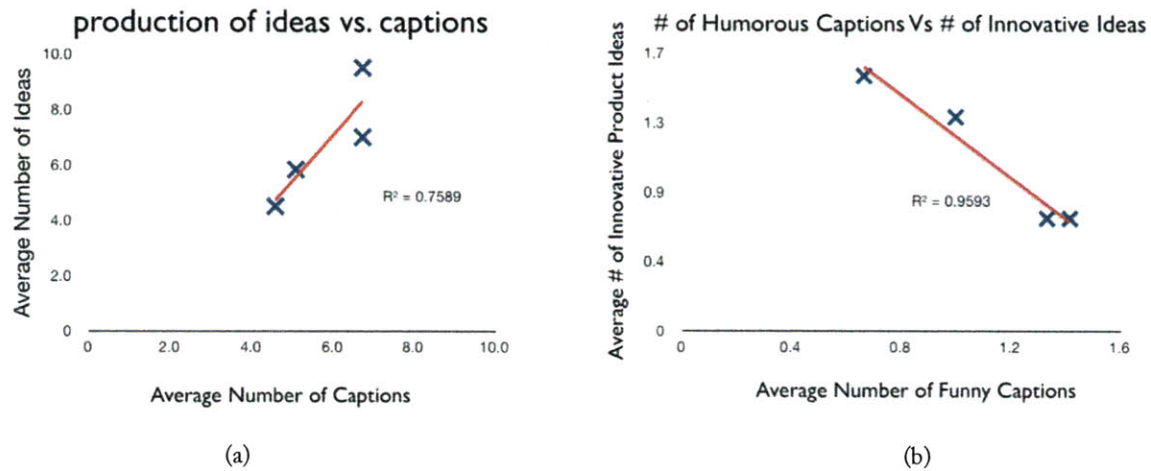
Figure 9. Quantity and Quality of Ideas vs. Quantity and Quality of Captions in Pretest

As expected, participants that produced a higher number of product ideas also produced a higher number of captions (r^2 = .76) as shown in Figure 9a. Somewhat unexpectedly and counter to the hypothesis of this thesis, participants that produced a higher number of funnier captions produced a lower number of innovative concepts (r^2 = .96) as shown in Figure 9b. This strong linear inverse relationship is very interesting. There are several possible explanations for this. The first being that this is such a small sample set of individuals with similar backgrounds and personalities that the data is not statistically significant.

Participants that produced a higher number of funny captions also produced a lower number of practical product ideas, novel product ideas and humorous product ideas as shown in Figure 10. Similar to the findings of Goodchilds, "creating humor and creating humorously are not the same. Nor does success at one imply the success at the other [6]." This trend, however, did not hold true in later testing with a larger sample size.
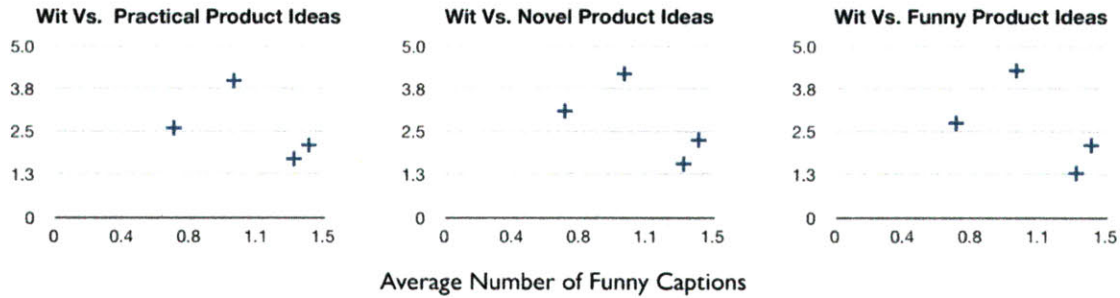
Figure 10. Average Number of Funny Captions vs. Practical, Novel and Funny Product Ideas (In Pretest)

As expected, the more product ideas generated, the more innovative product ideas were produced (r^2 = .5) as shown in Figure 11b. This agrees with the popular belief that quantity breeds quality in ideation [63, 64]. Rather unexpectedly, the pretest found that a higher production of captions resulted in lower number of funny captions (r^2 = .77) as shown in Figure 11a. This data conflicts with the findings of Derks and Hervas [112] as well as the data in our later study with a larger sample size.
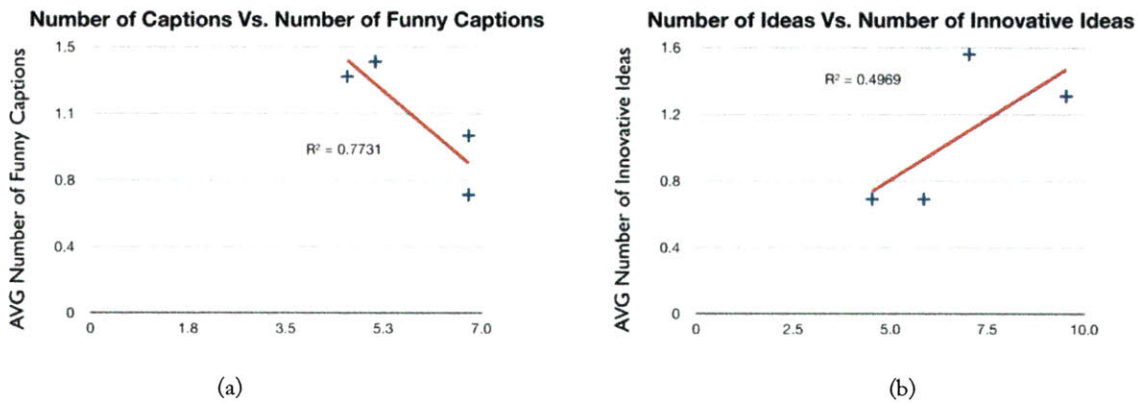
Figure 11. Quantity vs. Quality in both Captions and Ideas (In Pretest)

3.1.6 Remote Associates Test

Although this study focuses on a nominal product brainstorming as the test of applied creativity, we also wanted to explore the more traditional Remote Associates Test as a measure of creativity. The Remote Associates Test involves finding a connective link between sets of three seemingly unrelated words that have a mutually remote association.

In choosing which word sets and the time limit per word set, we referenced a paper that suggests different RAT tests for different challenge levels [143]. We chose thirty word sets that were of varied degrees of difficulty (see Appendix A). On a laptop, each word set was numbered and displayed for 15 seconds followed by a beep sound, which indicated that the next word set would appear. Subjects were asked to write the connective link on the blank line that corresponded with the number of the word set. Subjects were first given the example set: Cottage Swiss Cake. It was explained that "Cheese" is the connective link that can be used to make the phrases "Cottage Cheese," "Swiss Cheese," and "Cheese Cake." The connective link can come before or after the words shown.

A group of 24 students (13 male, 11 female, average age 20) were given the test. In this study, the RAT score is a number out of 30 potentially correct responses. The average score was 12.8 and the standard deviation was 4.2.

If we compare the number of correct scores on each question (with 24 subjects) with the scores from the original study [143] (with 100 subjects), the data is moderately-high correlated (r^2=.6). In other words, a similar percentage of subjects got each question correct between studies. This is shown in Figure 12
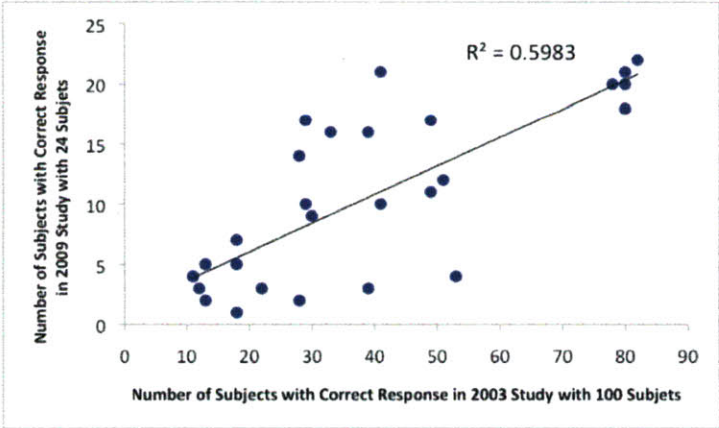
Figure 12. Comparison of Question Difficulty between RAT pretest and a Prior RAT study

3.2 Cartoon Caption and Product Brainstorming Experimental Setup

3.2.1 Participants

We wanted to determine if improvisational comedians, being well versed in humor production, are also skilled in product idea generation. We also wanted to see if professional product designers, being well versed in idea generation, are skilled in humor production. In addition to professionals, we tested MIT students as well as a mix of participants that are not designers, comedians or students. This mix of participants was not considered a control group, as it is not a statistically significant representation of the general population of the United States or even the greater Boston area. We will refer to this population as the "other" group, which includes participants that are not designers, comedians or MIT students.

(a)   (b)
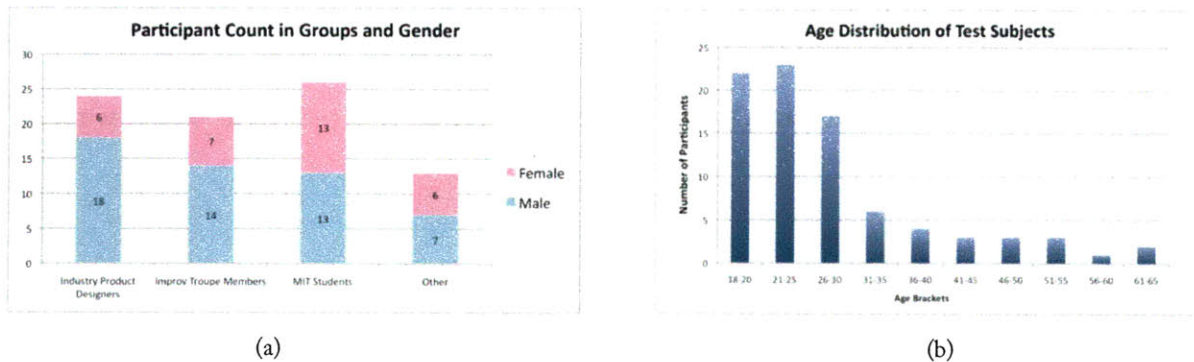
Figure 13. Age and Gender of Test Subjects

We administered this test to 84 participants (52 male, 32 female) in four interest groups: 24 product designers, 21 improvisational comedians, 26 MIT students, and 13 "other". The gender breakdown per group is shown in Figure 13a. The combined age range of the participants was between 18 and 63 years. The mean age was 28 and the median age was 23. The low average age was a result of the large number of MIT students and young designers and comedians. The age distribution is shown in Figure 13b.

The MIT students that volunteered to participate were all involved in one of three different product design-related courses (2.00b Toy Product Design, 2.009 Product Engineering Processes, or 2.97 DesignApalooza). The majority of these students were underclassmen and more specifically freshmen.

The product design industry volunteers were recruited from the MITRE Corporation in Bedford, MA and Hasbro Incorporated in Pawtucket, RI. These designers had anywhere between 1 and 38 years of experience in industry.

The improvisational comedian volunteers were recruited from ImprovBoston in Cambridge, MA and Road Kill Buffet, the MIT improvisation comedy troupe. These comedians had anywhere between 1 and 17 years of experience in the field.

The "random" group of individuals was recruited in different manners. Some are administrative assistants at MIT and some are family members and acquaintances of MIT students.

In the survey, the participants were asked questions related to their own assessment of wittiness and creativity. They were also asked questions on how often they practice the arts, do creative activities, make others laugh, and build things. As shown in

Figure 14, the majority of participants thought that they were creative (~86%) as well as witty (~81%).
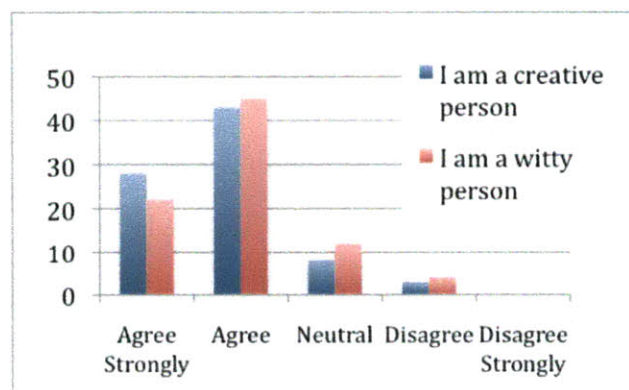
Figure 14. Participant Self-Assessment of Creativity, Wit, Etc.

3.2.2 Overview of the tests and forms

The test is composed of four parts. The first part is a consent form to participate in non-biomedical research. All participants received and signed this form prior to testing. A brief description of the tests was presented in the consent form.

The second and third parts of the test, the nominal product brainstorming and the cartoon caption test, alternated order in every test group. This was done to see if the order of testing plays a role in the creativity of the ideas or the humor of the captions. Studies have found that being exposed to humor prior to creativity tests increases your creative potential [57, 144, 145]. It was also found that being exposed to a random set of words prior to a brainstorm session also increased your creative output [146]. One could also argue that brainstorming before a humor test could increase your humor production abilities.

The final part of the test included a survey (see Appendix B and C) that asked questions regarding prior experience with idea generation and the arts, self-assessment of creative and humor abilities, assessment of the playfulness of the tests, and the satisfaction with the responses. There were also several questions related to factors that increase levels of dopamine (caffeine, sugar, movies, hugs, exercise, etc). Dopamine plays a significant role in creative thought and association making [122]. There was space for participants to write in comments on humor production and idea generation as well as define an innovative product in their own words.

Participants were given the test in familiar locations, typically in their work surroundings in a group at a large conference table. Tests were performed individually but sitting together as a group to save time on the part of the administrator and to make the participants more comfortable. An example of a testing environment is shown in Figure 15.

The entire test, including the briefing and consent form, the cartoon caption test, the nominal brainstorming test and the survey takes approximately 60 to 75 minutes.
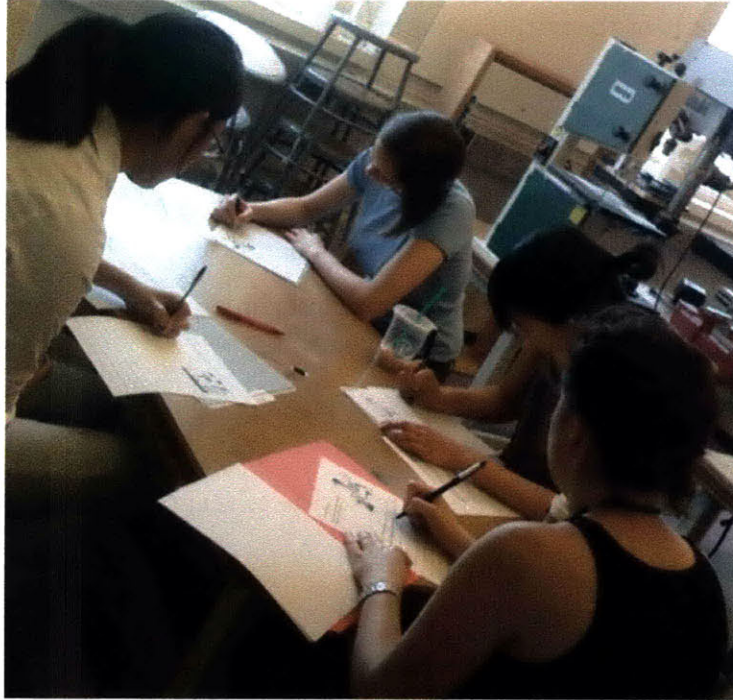
Figure 15. Example of a Test Environment with Four Participants

3.2.2.1 Cartoon Caption Test

Three cartoons chosen from the New Yorker Cartoon Caption Game were used for this test. These three cartoons, shown in Figure 7, were presented in random order for each participant. An example of a cartoon caption test sheet is in Appendix D. Each of the three test pages includes a caption-less cartoon and series of blank lines. Participants were given five minutes per cartoon. There were told the following:

> You will be given three caption-less cartoons in a random order.
>
> For each cartoon, please write as many humorous captions as you can in 5 minutes.
>
> I will say "time" at 5 minutes and you can finish anything that you started.
>
> Please write legibly.
>
> Please start each caption on a new line.
>
> Write all captions that come to mind even if you feel they are only mildly amusing.
>
> Do not talk to others.
>
> Label the order in which you receive the three cartoons.

3.2.2.2. Nominal Product Brainstorming Test

The three products chosen for the nominal product-brainstorming test were umbrellas, toasters, and toothbrushes. Each group of participants received a different order of these three products. Each brainstorm session lasted 12 minutes. Each participant was given a stack of blank legal paper and a black fine tipped Sharpie® permanent marker. The participants were told the following:

> You will be given three product themes, such as lamps or pens.
>
> For each theme, ideate as many innovative product concepts as you can in 12 minutes.
>
> I will say "time" at 12 minutes and you can finish anything that you started.
>
> Please sketch large and legibly.
>
> Drawing ability does not matter.
>
> Use as few words as possible, but give each idea a title.
>
> Each idea should be sketched on a new page in portrait orientation.
>
> Sketch all ideas that you feel are innovative in some way.
>
> Do not talk to others

3.3 Method of Evaluation of Captions and Product Ideas

3.3.1 Choosing Evaluation Metrics for Cartoon Captions

Evaluating the cartoon captions is relatively simple compared to the product ideas. We will use a format similar to the pretest and the prior caption studies of Table 4. Participants' unique responses were counted for fluency scores. The captions were then digitized with grammar corrected. Judges then evaluate the funniness of the captions on a scale. All prior studies have a panel of judges read through all captions. These panels range in size from 2 to 13. For this study we used a website that is open to the general public to visit and rate random captions. We feel that this method is best for reducing rater fatigue and getting a wide variety of senses of humor to evaluate the captions. Humor like creativity is subjective and, following the subjective assessment method of Amabile [68], something is funny if an appropriate panel of reviewers thinks it is funny.

Rating of funniness should be based on a set scale as opposed to a ranking, a decibel laugh measure, or a binary funny/not funny [21]. One prior study used a binary funny/not funny rating [11], but other prior studies use scales of 4 points and greater. If reviewers are going to read hundreds of captions, a large point scale is only effective if each point value comes with a qualitative

description. We feel that a descriptive three-point scale (2=funny, 1=somewhat/moderately funny, 0=not funny) similar to Turner's 4-point scale [110] is an effective way of quickly and consistently judging captions. Captions should be randomized for each reviewer. Reviewers should do the rating individually where they cannot influence other reviewers. 12 people reviewed each caption. The largest panel size used in any prior cartoon caption test was 13.

The scores for each caption were averaged and a caption was considered "funny" if the average score is above a 1.0. A participant's humor score can be measured several ways, but we will focus on one measure: a total count of 2s for all captions produced by that participant. We will use counts opposed to averages or percentages, as they do not penalize participants when they produce many unfunny captions along with funny captions.

3.3.2 Choosing Evaluation Metrics for Product Ideas

In having people review the ideas for the pretest there was much confusion on the meanings of usefulness, practicality, feasibility and novelty and whether these criteria actually determined innovativeness or creativity.

In choosing appropriate evaluation metrics, we referenced the prior studies of Table 2. Dean et al [74] claims that there are 80 studies that deal with reviewing ideas from 1990-2006. Although his classification, based on MacCrimmmon and Wager [60] is thorough, it is tailored to evaluate ideas for themes that are outside the realm of products and design such as "how to improve the business department" or "how to make Tucson more of a tourist location." Dean suggests reviewing ideas by the following classification: Novelty composed of Originality and Paradigm Relatedness, Workability composed of Acceptability and Implementability, Relevance composed of Applicability and Effectiveness, and Specificity composed of Implicational Explicitness, Completeness, and Clarity.

Horn and Salvendy [79] and Besemer and Treffinger [7] deal specifically with metrics for reviewing products and designs and not simply solutions to hypothetical problems. The one concern with their metrics is that they deal with products that already exist and are on the market. They also only review a few designs as opposed to hundreds of concepts and so their metrics can be long and detailed.

There are no existing metrics specific to conceptual product design review, so we must draw upon appropriate metrics from prior studies. We evaluated ideas based on quantity as well as five qualitative metrics: Creative, Novel, Useful, Product Worthy, and Clear. Descriptions of these metrics are as follows:

Creative – *Using the judge's own subjective definition of creativity* [68]. Following in the subjective assessment style of Amabile and Christiaans, we decided to ask the reviewers to rate the ideas on several metrics, one of which is creativity as defined by the reviewer [37, 68]. We can then assess which metrics correlate with the subjectively rated creative products.

We also decided to have reviewers subjectively rate on "creativity" of the products rather than "innovativeness" of the products. Even though the definition of "creativity" is subjective, we believe that majority of people have a better internalized meaning of "creativity" as compared to "innovative." It is possible to then take a more traditional objective definition of "innovative" such as the combination of knowledge in original and non-obvious valued new products, processes or services [29] which would equate to a formula such as "novel" plus "useful" plus "product worthy."

Novel - *The concept is original [76] and uncommon* [79]. A gun that shoots out celery may not be considered useful or practical, but it should be considered novel. According to the multi-attribute definitions of creativity, novelty does not always imply creative. Take the celery gun example and imagine if the prompt was to ideate new toothbrush concepts. This concept would be considered original and uncommon, but would not make any sense in the realm of toothbrushes. As Mednick explains "7,363,474 is quite an original answer to the problem 'how much is 12+12'" but it would not be considered creative [3].

Useful – *The concept has practical applications in the product theme*. This is perhaps the most controversial category for rating concepts. MacCrimmon and Wagner [60] and Dean et al [74] would call this category "Relevance" meaning applicable the problem or topic at hand with an effective solution. This wording is more appropriate for structured prompts as they have solutions. Horn and Salvendy [79] call this category "Importance" defined as relevant

and significant. Besemer and Treffinger [7] and Bessemer and O'Quin [77] would call this category "Resolution" defined as the correctness of the solution; it is logical, useful and valuable. In this research study we decided to separate "value" from "use" as products can have value without being of practical use.

Product Worthy – *The concept is both feasible and marketable.* Feasibility can be described as the ability of the idea to be physically made into a product. It takes into account: technology, cost, safety, and manufacturability. Marketable takes into account the social and cultural variables and if people would want to purchase the product (is it too much work to use, is it too large to store, would it be too expensive for the intended users). This category is termed "Workability" by MacCrimmon and Wagner [60] and Dean et al [74], and they define it as acceptability and implementability. This category is not addressed by Horn and Salvendy [79] or Besemer and O'Quin [77] as their studies review existing products. Using our definitions, there are concepts that are product worthy but are not useful (of practical application). These items would be considered "novelty" items, aesthetic innovation, or Chindogu.

Product Worthy may be hard to determine for the layman, as there are many different technologies that are unknown to the general public and some ideas require a basic understanding of physics. The layman also may not have a good sense of what is marketable.

Clear – *The concept is well communicated.* This metric maps to what Dean et al [74] and MacCrimmon and Wagner [60] term "Specificity" meaning clear, complete and explicit. As we are dealing with brainstorming, it is not expected that ideas should be elaborate and complete, however we are expecting ideas to be clear and detailed enough that the judges are able to understand the concept. This category also relates to the Besemer an O'Quin [77] category of "Synthesis" which includes organic, well-crafted, and elegant and once included understandable [76]. Clarity is important in ideation as well as in humor production. Unlike "creative", "novel", "useful", and "product worthy," "clarity" is independent of the concept and based solely on the participant's presentation.

We were also interested in adding the metrics of "humorous," "playful," "desirable," and "I would buy this," but given the quantity of product ideas to review, additional metrics could contribute to review fatigue.

If the multi-attribute based definitions of creativity are correct, we should see a strong correlation between the subjective assessment of creativity and a combination of high ratings on novel and useful and/or product worthy. If the novelty based definitions of creativity are correct, we should only see a strong correlation between the subjective assessment of creativity and high novelty ratings.

Reviewers will be asked to rate each idea on these five metrics: Creative, Clear, Novel, Useful, Product Worthy. To make these metrics easy to understand, we included a short explanation. Each metric was rated on a 3-point Likert scale (2 = yes, 1 = somewhat, 0 = no).

Creative – The concept is creative.

Clear – The concept is well communicated.

Novel – The concept is original and uncommon.

Useful – The concept has practical applications in the theme.

Product Worthy – The concept is feasible and marketable.

When evaluating a set of ideas (*i.e.* the creativity of a participant) we believe that a total count of 2s is the most equitable measure to ensure that good ideas are not devalued by a large number of bad ideas [37]. A count of good ideas (*e.g.* number of ideas with a score over 1.0) is also an acceptable means of scoring the individual participants, however, the count of 2s gives a greater depth and finer resolution.

3.3.3 Online Review and Judges

To evaluate the product concepts and the jokes, we chose to use an online website approach as opposed to a physical review form. An online review has several benefits. The reviewers can be located in many different locations and thus we can get a better general population sample. The reviewers can do the rating at their own convenience and in a comfortable setting. The data can be

reviewed by hundreds of people as opposed to a select panel of individuals. The review data is also easy to collect.

The prior studies that involved ratings of captions (Table 4) used a small group of individuals as judges of humor. These groups ranged in size from 2 up to 13 members. As we have found in our testing, people find different things funny and a review panel of two is not sufficient to cover the range of styles humor appreciation. These prior studies also used humor scales that range from binary (funny/not funny) to upwards of 11 points. We feel that a scale closer to binary is the only appropriate measure when dealing with hundreds of captions.

We used Amazon Mechanical Turk as the means of collecting reviews. Amazon Mechanical Turk is a website (http://www.mturk.com) that allows any user to post tasks for any other user to complete. The tasks are termed "HITs" or "Human Intelligence Tasks," the person posting the task is termed a "requester", and the people completing the tasks are termed "workers." Requesters post HITs and a monetary payment for completing each HIT.
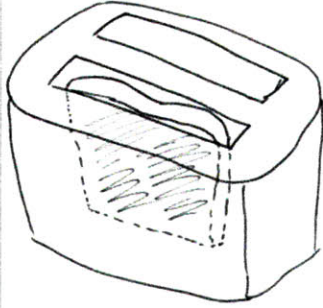
In this study, each worker was paid $0.15 to rate 20 randomly selected product ideas or $0.15 to rate 24 randomly selected cartoon captions. Each product idea was presented as a scan of the original sketch alongside radio buttons for rating as shown in Figure 17. Each set of 24 captions was listed below the corresponding cartoon image as shown in Figure 18. Twelve different workers out of a pool of 397 (an average age of 34.4, approximately 38% male) rated each of 545 toaster ideas, 627 umbrellas ideas and 595 toothbrush ideas. Twelve different workers out of a pool of 437 (an average age of 34.3, approximately 44% male) rated each of 1398 captions. All workers were located in the United States. Workers were asked to fill in a few survey questions prior to testing as shown in Figure 16.

Please provide the personal information and then evaluate the following conceptual toaster concepts

1. What is your gender?

○ Male
○ Female

2. What is your age?

_____

3. Which of the following best describes your highest achieved education level?

[Some High School   ▼]

4. Are you affiliated with the product design industry?

○ Yes
○ No

Figure 16. Example of Online Survey Prior to Testing

S2250000010.jpg

The concept is Creative.
○ Yes   ○ Somewhat   ○ No
The concept is Novel (uncommon and original).
○ Yes   ○ Somewhat   ○ No
The concept is Useful (practical applications in the theme).
○ Yes   ○ Somewhat   ○ No
The concept is Product Worthy (feasible and marketable).
○ Yes   ○ Somewhat   ○ No
The concept is Clear (well communicated).
○ Yes   ○ Somewhat   ○ No

Submit

Figure 17. Example of Online Product Review Form

D222	The company is underwater... and we called for you.	○ Funny ○ Somewhat Funny ○ Not Funny
I208	We're expecting the stock market to go to the toilet and may need you for assistance.	○ Funny ○ Somewhat Funny ○ Not Funny
	Boss: You are fired / Employee: ...	
S226	(good thing I can hardly hear anythign in this helmet)	○ Funny ○ Somewhat Funny ○ Not Funny
I216	Wait, if you're here... where the hell is Johnson?	○ Funny ○ Somewhat Funny ○ Not Funny
I215	Who you calling a beekeeper?	○ Funny ○ Somewhat Funny ○ Not Funny
D216	I said "a *nice* suit" not "a space suit"	○ Funny ○ Somewhat Funny ○ Not Funny

Figure 18. Example of Online Cartoon Caption Review Form

3.4 Design of an Improvisational Training Workshop

A recent conference paper on improvisation and brainstorming concluded that empirical studies are needed to better understand the relationships between these fields [43]. To test the effect of improvisational training on idea generation, a workshop was designed based on short form improvisation. This workshop is intended to precede an idea generation session to stimulate associative thinking.

3.4.1 Choosing Appropriate Short Form Games and Exercises

The following are a list of short form improvisation games that are most related to training for prolific non-obvious associations. Many of these games are similar to those suggested and described by Gerber [43] others are commonly used in introductory improvisational theatre classes and described online in the Improv Encyclopedia [147]. There are a few additional games in the beginning of this set that are less related to associations, but good for introductions and participation.

The entire workshop lasts between 60-90 minutes. The length of each game varies depending on the time it takes for everyone to fully participate.

Name Thumper – The main goal of this game is an introduction. All players begin in a circle. Each player states their name and does a unique motion. Then going around the circle, all players say each name and do each motion. One player begins by saying their name and doing their motion and then saying another players' name and doing their motion. That player then continues until everyone knows each other's names.

Samurai – The goal of this game is a warm up and to get players to be alert. All players begin in a circle. One player is the Samurai - she starts by lifting her Katana (sword) and yelling "Wah." She keeps her sword up in the air, and her two neighbors "slaughter" her, by swinging their swords into her sides, simultaneously yelling "Wah." When the neighbors retract their swords, the Samurai lowers her sword making eye contact with another player and yelling "Wah," who then becomes Samurai. This sequence continues and gradually becomes faster.

76

Red Ball – This game allows players to start making associations related to a simple item (i.e. a ball). To begin, players are in a circle. An imaginary red ball is tossed between players. The ball can be transformed into different objects but maintains some connection to a ball. Players have a bit of time to think of possible options and given the simplicity of a ball, it can be taken in many different directions. The receiving player must pay attention to understand what form the ball is taking on. To pass the ball, the player looks at another player and says "red ball." The receiver then says "red ball" to indicate that they are ready and aware of what form the ball has taken on. Other color balls can be introduced. This game is almost a 3-D version of the TTCT "Circles Test" in which a subject is asked to sketch as many different objects as possible using a set of blank circles.

Word Ball – The goal of this game is to make quick verbal associations. All players begin in a circle. One player starts by tossing a word to another player. The receiver thinks of the first word that comes to mind, and throws his word to yet another player. This game gets players to take a concept and quickly make an association on it. Players do not necessarily have to make non-obvious connections.

Story Spine – This game is a good example of the "yes and…" rule, where the goal is to listen to what someone says and build on it.

In this game, players tell a story one line at a time following the structure:

- Once upon a time ...
- And every day ...
- And then one day ...
- And because of that ...
- And because of that ...
- Until Finally…
- And ever since that day ...

By providing a general story template, players are more comfortable building original content.

Yes Let's – The goal of this game is to get players to break out of the circle and become comfortable stating ridiculous ideas. One player begins by saying "Hey everyone! Let`s ..." filling in an action for everyone to do, such as "jump out of a plane" or "be in a slow motion fight scene". Then everyone loudly agrees "Yes, Let's!" and begin said activity. This continues until another player shouts "Hey everyone…Let's." This game is played until everyone has suggested something.

Emotion + Item - This game gets players quickly making associations on random items suggested by other players. Players are in a line. An object is suggested as well as an emotion and every player steps forward one at a time and declares why they feel that way about that object. Next, one player steps forward and declares why they feel someway about an object X in the form: I ____ X because Y. Any other player steps forward and declares how they feel about some part of Y in terms of Z. This continues taking the reason for the emotion and turning it in to the subject of the following statement.

Freeze Tag – This game gets players to make associations based on body positions and actions. It encourages looking at something and seeing it in many different perspectives. Two players take the stage and pose. Another player tags out one of the players on stage and makes a new pose that could potentially take a scene in a new direction. Players continue to tag each other out making various static scenes. This can evolve to each player saying a line when they change the scene. The next step would be having players improvise a scene and at any point another player can call freeze. This player then tags out one of the two players on stage and takes his place. Both players then start a new scene, justifying their positions.

Two other games that are based on associations, but were not included in this workshop are Ding and Props.

Ding – A scene is played. Whenever the MC rings a bell (or yells `Ding`), the player that is doing something/saying something needs to say/do something else.

Props - In this game the players need to come up with an original/funny use of a prop. They cannot use the prop for its intended use. This game is very similar to the "Unusual Uses Test" or "Alternative Uses" test in which the subject is asked to think of as many unconventional uses for each of a number of objects

3.4.2 Procedure, Participants and Review

The procedure for administering the nominal brainstorming test (including instructions, materials, time) remained the same as presented in 3.2.2.2. Nominal Product Brainstorming Test. However, no cartoon caption test was administered and ideas were only counted and not rated. On the first day, the subjects were asked to do a nominal brainstorming session on one product theme

(toothbrushes).   On the second day, the subjects participated in an hour-long improvisation workshop followed by a second nominal brainstorming session on a different product theme (umbrellas).

In this study, only the quantity of ideas was measured for each subject, as we were mostly interested in how fluency of a brainstorming session is influenced by improvisation theatre training.

4. Results

4.1 Interrelationships between Product Theme Fluency

In comparing individual fluency between product ideation themes (Figure 19) there appears to be high correlation between all three products: toothbrush and toaster (r^2=.69), toothbrush and umbrella (r^2=.75), and umbrella and toaster (r^2=.67). This implies that when dealing with blue-sky idea generation in a common household product theme, fluency does not depend on the specific product theme. The stronger correlation between toothbrush and umbrella may be related to the fact that both of these products are extensions of the arm.
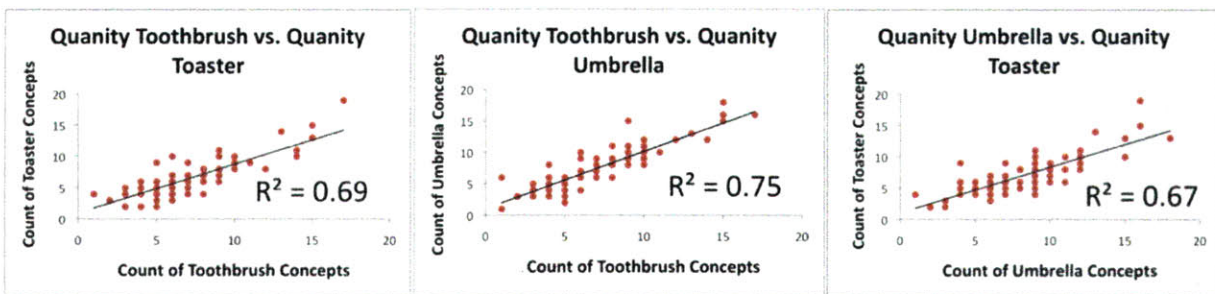
Figure 19. Correlation of Ideation Fluency between Product Themes

4.1 Interrelationships between Metrics

In this study we had reviewers evaluate several metrics including: creativity, novelty, product worth, usefulness, and clarity. Table 8 shows the interrelationships between these metric rating.

Table 8. Interrelationships between Product Metric Ratings

	Clear	Creative	Novel	Product-worthy
Creative	.32			
Novel	.24	.80		
Product-worthy	.43	.14	.07	
Useful	.42	.16	.08	.86

Useful and product-worthy (r^2=.86) as well as creative and novel (r^2=.80) are practically indistinguishable to reviewers. Note the low correlation between creative and useful scores (r^2=.14). Clarity appears to be moderately correlated to product-worthy and useful ratings, and minimally correlated to creative and novel ratings. One can notice greater effects of clarity on

81

creativity when comparing a set of similar ideas that are sketched with different levels of clarity. This is discussed in later sections.

Figure 20, Figure 21, and Figure 22 plot the most relevant inter-metric relationships.
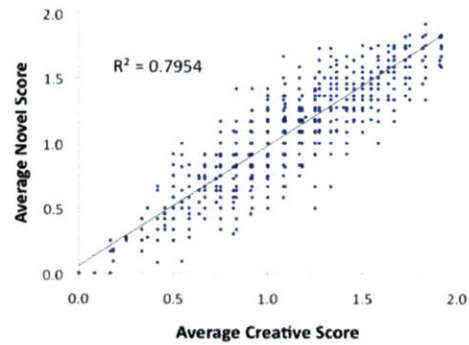
Figure 20. Creative Scores vs. Novel Scores for Each Idea
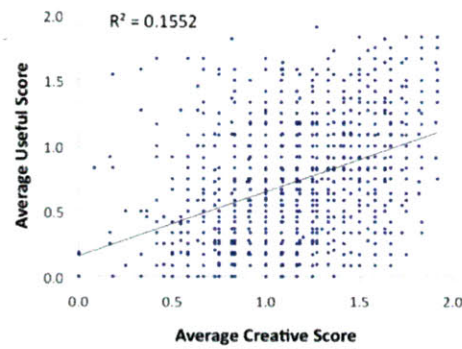
Figure 21. Creative Scores vs. Useful Scores for Each Idea
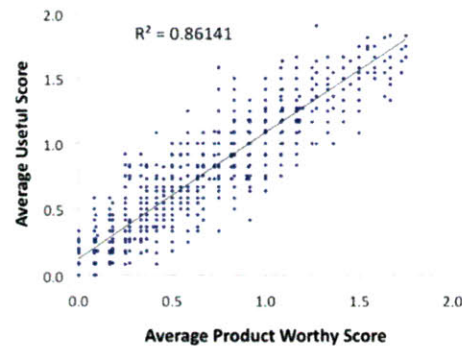
Figure 22. Product Worthy Scores vs. Useful Scores for Each Idea

4.2 Demographic Comparison

Improvisational comedians on average produced 17% more captions and 20% more product ideas than product designers. They also produced approximately 28% more captions and 44% more product ideas than the "other" group. As far as quality of the output, the improvisational comedians on average had approximately 32% higher funny caption scores (2s) and 21% higher creative product scores (2s) than the product designers. When comparing to the "other" group, the improvisational comedians had 42% higher caption scores and 48% higher creative product scores. These demographic comparisons can be seen in Figure 23.

(a)   (b)
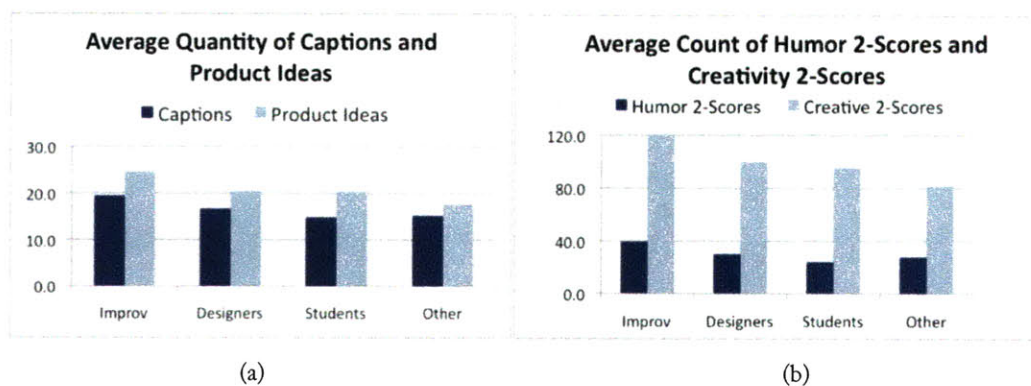
Figure 23. Comparison of Average Quantity and Scores by Interest Group

There was very little difference in quantity of captions and quantity of ideas between genders as shown in Figure 24. As for count of 2s, females seem to have slightly lower humor scores and slightly higher creativity scores, but these values may be negligible. The subjects included 52 males and 32 females.
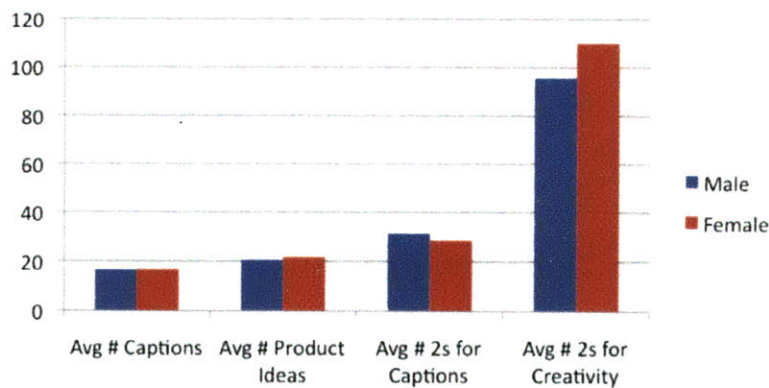
Figure 24. Gender on Creativity and Humor Production

4.3 Quantity of Ideas

Quantity of ideas was highly correlated with overall individual creativity scores (r^2=.82) as well as quantity of creative ideas (*i.e.* ideas with average creative scores over 1.0) (r^2=.64). It may be argued that the high creative scores is an artifact of having a lot of ideas, however, the useful score is only minimally correlated with quantity of ideas (r^2=.38). The quantity of ideas compared to individual creative and useful scores are shown in Figure 25 and Figure 26 respectively.

Figure 25. Overall Quantity of Ideas vs. Creativity Score per Subject

Figure 26. Overall Quantity of Ideas vs. Useful Score per Subject

If we plot the quantity of ideas to the total average creativity score per subject, there is a negligible negative/no correlation (r^2=.01). This, however, is not an accurate measure as subjects are penalized for having bad ideas mixed with good ideas. 10 good ideas out of 100 is better than 1 good idea

84

out of 2, even though the former has a much smaller percentage of good ideas. It is the number of good ideas that matter, not the number of bad ideas.

Quantity of product ideas was highly correlated with overall individual creativity scores as a count of 2s (r^2=.82). Quantity of captions was highly correlated with overall individual humor scores as a count of 2s (r^2=.64). These "quantity" *vs.* "quality" graphs are shown in Figure 27. A few studies have found similar results in both idea generation [29, 40] and in cartoon caption tests [24].
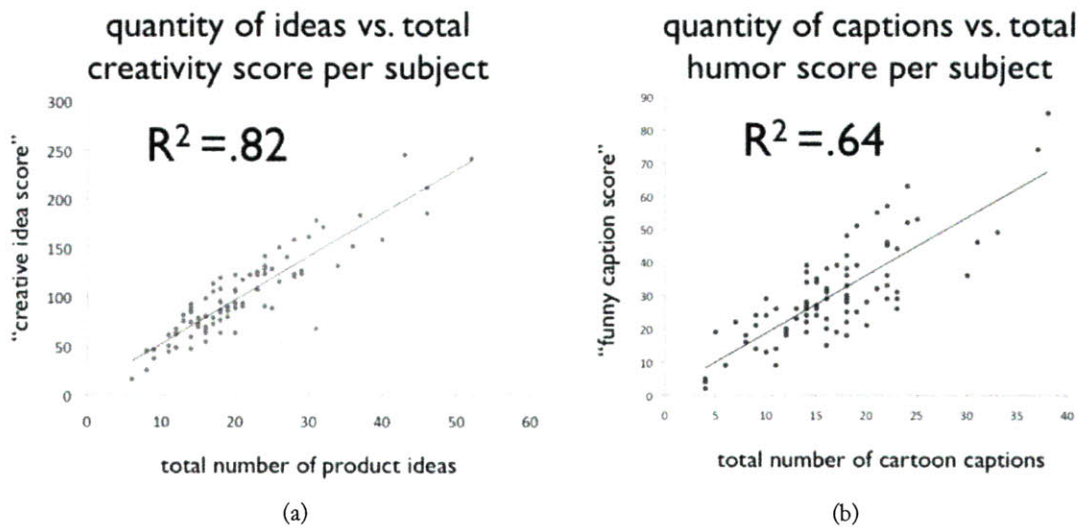
Figure 27. Quantity vs. Quality for Product Ideas and Cartoon Captions

In Figure 27a and Figure 27b, there seems be a small cluster of individuals that are highly prolific in both humor and idea generation. These individuals are to the upper right of the diagonal bisectors of Figure 28a and Figure 28b. Interestingly, 8/10 of the individuals located in this group for product ideation (Figure 28a) are also in this group for the cartoon caption test (Figure 28b). This suggests that highly prolific generators are highly prolific in different domains.

quantity of ideas vs. total creativity score per subject

quantity of captions vs. total humor score per subject

(a)   (b)
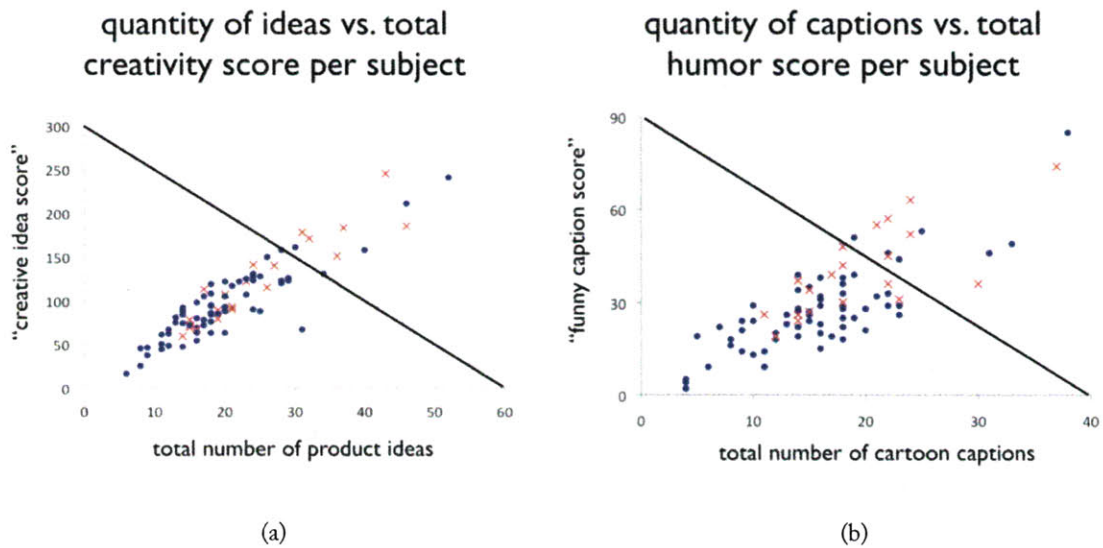
Figure 28. Improvisational Comedian Subject Placement (x) and Highly Prolific Generators

In Figure 28a and Figure 28b, the improvisational comedian subjects are labeled as Xs. Over half of the highly prolific generators are improvisational comedians. The other few subjects in this highly prolific range are professional product designers and one student (with a comedy writing background).

We were interested in whether the order of testing would affect the output. In testing, each group of subjects alternated order of the nominal product brainstorming and the cartoon caption test. Overall, 39/84 subjects received the caption test prior to the nominal brainstorming and 44/84 received the tests in the opposite order. Studies have found that being exposed to humor (as an observer) prior to creativity tests increases your creative potential [57, 144, 145]. We found that subjects that received the nominal brainstorming prior to the cartoon caption test produced approximately 22% more product ideas and 14% more captions as shown in Figure 29. The improvisational comedians had the most even distribution of subjects as far as order of testing with 11 subjects taking the caption test first and 10 taking the brainstorming first. In this subgroup, subjects that received the nominal brainstorming prior to the cartoon caption test produced approximately 38% more product ideas and 21% more captions. It is possible that taking the cartoon caption test prior to the nominal brainstorming fatigues the subjects both physically and mentally. This observation is in opposition to our hypothesis.
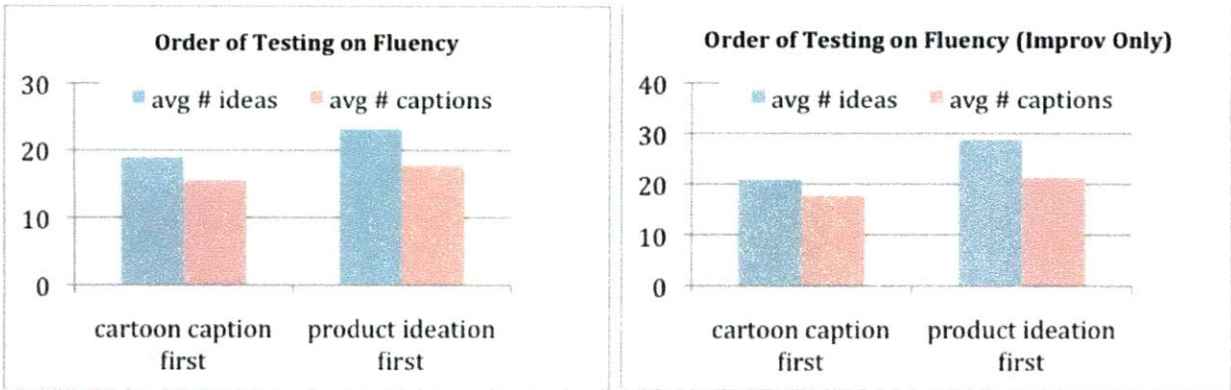
Figure 29. Order of Testing on Fluency of Output

4.4 Interviews and RAT with Highly Prolific Subjects

To better understand the thought process of the highly fluent idea generators, we interviewed 5 of the 10 subjects that fell to the right of the line in Figure 28a. 4/5 of these subjects had formal improvisational comedy training. The fifth was a comic strip artist for several years. The four subjects with improvisational comedy training also participated in 10+ years of improvisation in other forms including: jazz piano, jazz drums, Irish dance, and singing. Perhaps the improvisational ability or drive translates between mediums.

We administered a RAT to these five individuals (average age 22, 2 m 3 f) to see how they compare to the pretest of 24 students (avg age 20, 13 m, 11 f). The average RAT score for the five fluent idea generators was 15.6 with a standard deviation of 2.6. The average score of the 24 students in the pretest was 12.8 with a standard deviation of 4.2. This is shown in Figure 30.
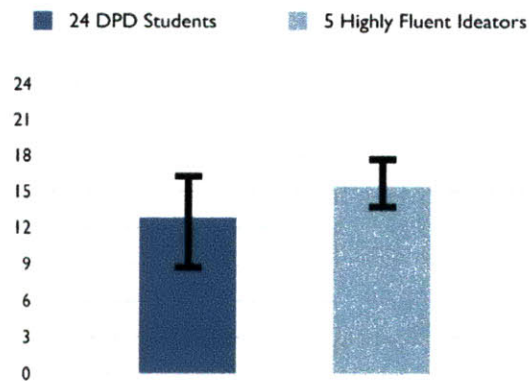
Figure 30. Average RAT Scores of Highly Fluent Ideators and a Group of Students

The fluent idea generators appear to score slightly higher than pretest subjects. It is debatable whether this test is an actual measure of creativity as there is a "correct" answer to each question. It is also purely verbal when compared to the product brainstorming which is both spatial and verbal.

In interviews following the RAT, these subjects were asked how they come up with ideas. Most responses involved making associations: "mash up different concepts," "change product attributes," "metaphor connections," "add fun features," "think about strange contexts."

The subjects with improvisation experience were also asked how their training has influenced their life. Reponses tended to reflect positive qualities in a creative designer: "I say all ideas that come to mind," "I don't over think," "I anticipate what may happen," "I actively think about alternatives," " I am always prepared for the unexpected," "I am comfortable talking to people," "I always build on ideas," "I handle conflicts," "I respond quickly with input".

4.5 Number of Reviewers

In this study we had a set of twelve people rate each product idea and each caption. Although this is a relatively large set of reviewers when compared to prior studies, we wanted to ensure that twelve reviewers is a good representative sample of people in the United States that are using Amazon Mechanical Turk.

We chose four captions from the "Spacesuit in Office" themed cartoon test that covered a range of average humor scores. We ran a second review on these four captions with 100 reviewers. The average age and gender breakdown between the tests was similar.

 For each caption, we combined the 112 scores and plotted the difference between the mean of 1000 random samples of 100 and the mean of 1000 random samples of 12. We then compared the original experimental difference to these histograms.

From this bootstrap analysis (Figure 31), it appears that there is very little difference between the average scores of twelve reviewers and 100 reviewers in all four of these examples. This is shown in the four histograms of Figure 31.
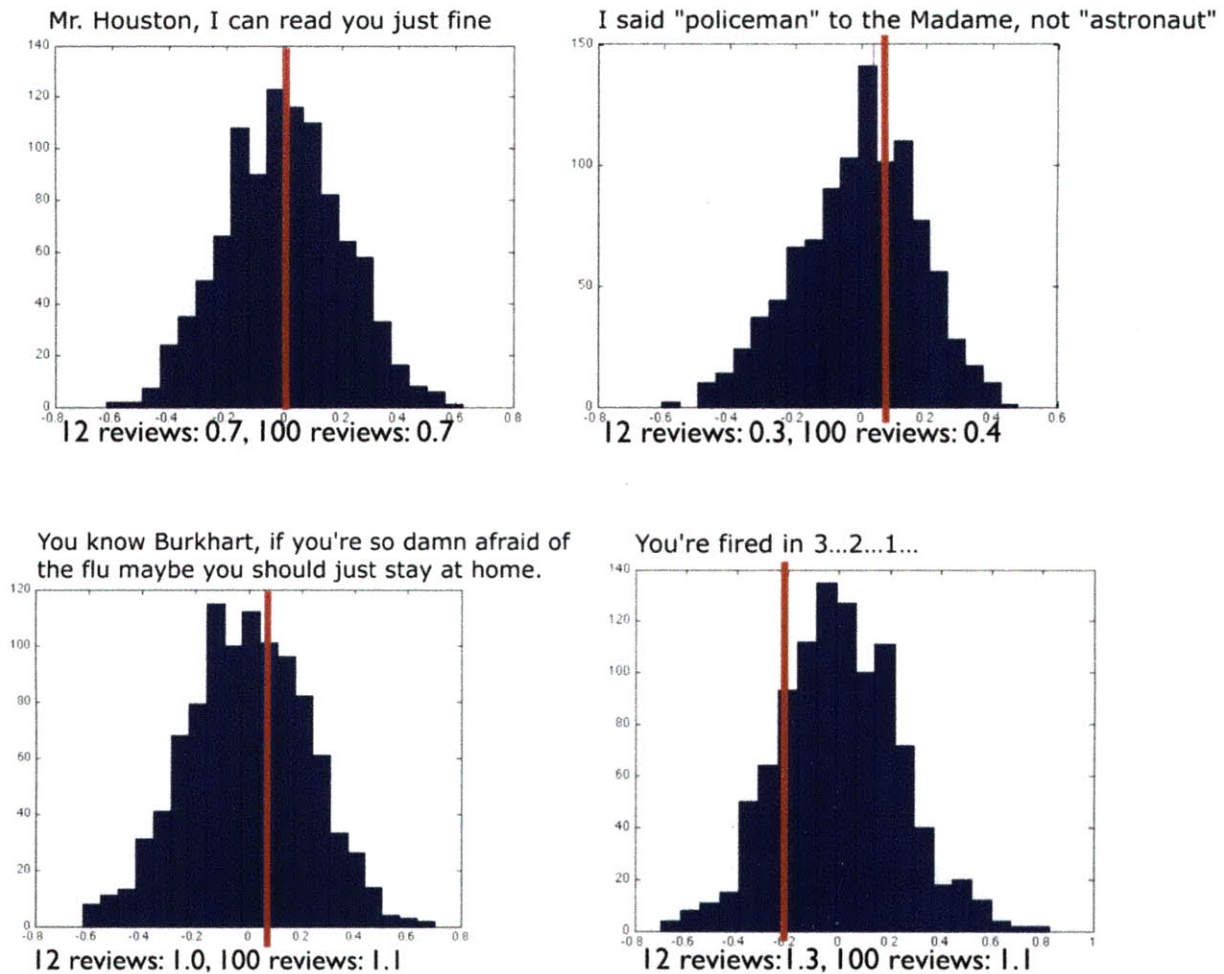
Figure 31. Bootstrapping Data Showing Difference Between 12 Reviewers and 100 Reviewers for 4 Cartoon Captions

4.6 Self Assessment Accuracy

As part of the survey, subjects were asked to respond to the statements "I am a creative person" and "I am a witty person" on a Likert scale of 1 (Strongly Disagree) to 3 (Neutral) to 5 (Strongly Agree). None of the participants used the 1 (Strongly Disagree). The creativity and humor scores were averaged for each Likert scale group (e.g. those that strongly agree they are creative). Figure 32a and Figure 32b show that subjects that thought they were more creative or witty had higher average creativity or humor scores respectively. This implies that people are good at judging their own creativity and humor abilities (as perceived by others).
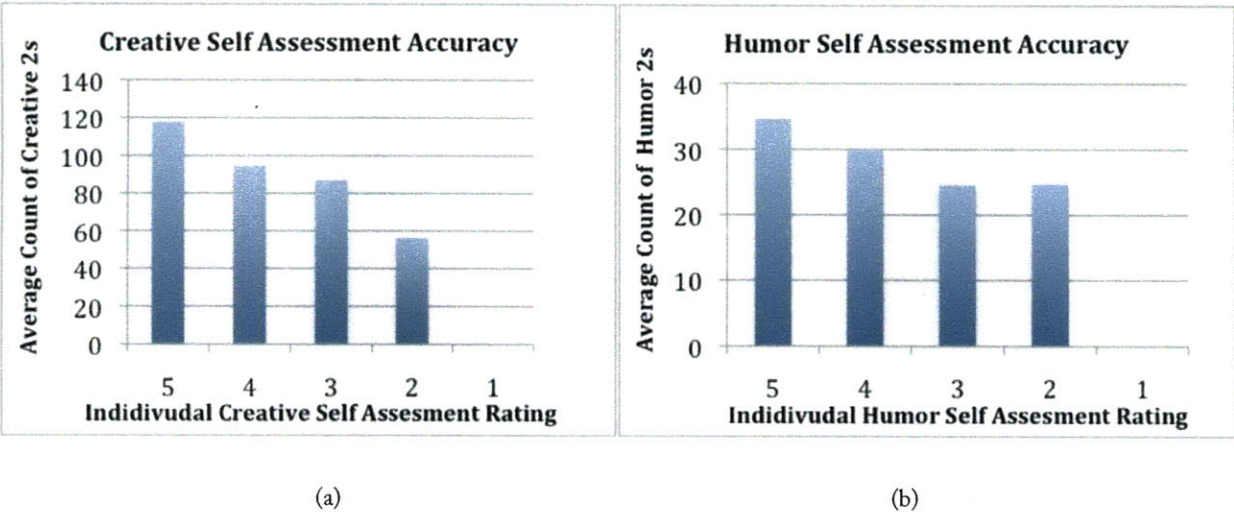
Figure 32. Self-Assessment of Creativity and Wit vs. Reviewer Assessment

4.7 Improvisation Workshop Effect on Fluency

Following this cartoon caption / nominal brainstorming test, a separate study was performed to determine the affect of an improvisational workshop on idea generation fluency.

Students were recruited for this study knowing that it entailed an improvisation workshop and product brainstorming. The subjects for this group were 11 students (7 female, 4 male).

In this test, the group was asked to brainstorm toothbrush concepts on the first day. The following day they received an hour-long improvisation workshop followed by a nominal brainstorming on umbrellas.

Given the same time, product theme, and prompt, the 11 MIT students that signed up for the improvisation workshop/test had twice as many ideas on average *prior* to improvisation training, than the group of 26 MIT students from the original study. Perhaps the students that are willing to participate in an improvisation workshop are more open minded, willing to think differently and thus are going to be more prolific generators.

After the workshop, these 11 students increased in idea fluency by approximately 37%. This suggests that even highly fluent generators can become even more fluent with training.

11 students
11.4 Ideas/subject Avg. (prior to workshop)

22 students
6.9 Ideas/subject Avg. (no workshop)

11 students
15.5 Ideas/subject Avg. (after workshop)

22 students
7.15 Ideas/subject Avg. (no workshop)
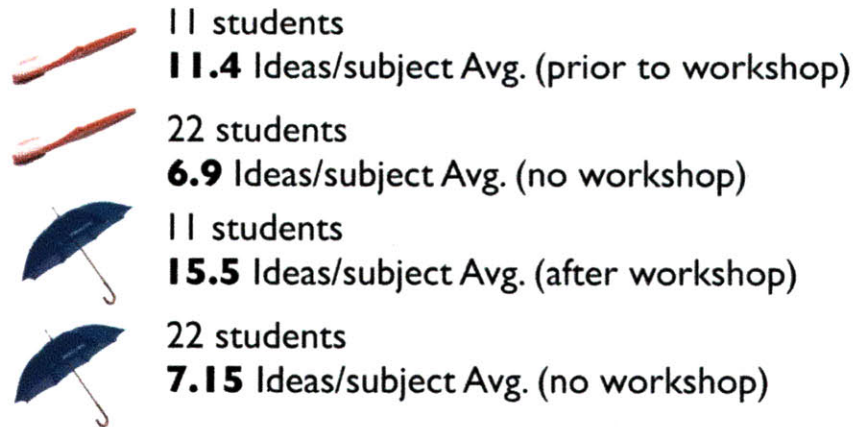
Figure 33. Pretest Summary of the Affect of Improvisational Workshop on Idea Fluency

5. Discussion

5.1 Quantity of Ideas

As quantity of ideas is highly correlated with creativity of ideas ($r^2=.82$), it is reasonable to measure a individual's creativity simply by fluency. This, however, does not accurately represent that individual's ability to produce useful or feasible ideas.

It could be argued that individuals that produce a lot of ideas are better at divergent or associative thinking, which is often related to creative thought process [32, 33]. As more associations are made, the probability of reaching a creative idea increases [22]. One could also argue that individuals that are uninhibited will edit thoughts less, produce many ideas and their output should be less restrained and thus more creative. Oppositely, individuals that are logical thinkers may come up with more useful concepts, but will also be restrictive in their thought process producing fewer concepts in total.

The high correlations of quantity of ideas to creativity of ideas as well as quantity of captions to humor of captions supports the quote by Linus Pauling that "the best way to have a good idea is to have a lot of ideas."

Almost all prior cartoon caption studies ask the participants to produce one caption per cartoon. We believe that this does not capture an individual's true humor production ability.

5.2 Rating Product Worth

We defined product-worthy as both feasible and marketable. These elements can be mutually exclusive which made this category difficult for reviewers. Laymen reviewers had difficulty rating marketability. Several ideas were rated as not marketable, when the idea was for an existing product. An example of this is the Miracle Toaster shown in Figure 34a, where 12/12 reviewers rated it as not product-worthy when in fact it is currently a product on the market sold by Fred Inc. designed by Jason Amendolara, shown in Figure 34b.

<center>(a)   (b)</center>

Figure 34. Toaster that Produces Holy Images on Toast (http://www.worldwidefred.com/holytoast.htm, 2006)

Laymen are also not appropriate raters of feasibility. A good percentage of reviewers rated a battery-powered toaster, a hand-cranked toaster and a cardboard box solar toaster as product-worthy.

As marketability is dependant on a variety of factors external from the concept itself, we suggest simplifying the category of product-worthy to feasible with the caveat that it is to be rated by expert judges (engineers or designers).

5.3 Concept Clarity and Creativity

The general correlation between clarity and creativity scores was low-moderate (r^2=.32). However, with the large number of ideas generated, there were many ideas that were reoccurring, presented by different participants. Using these sets of ideas, we can better see the effect of clarity of sketch on perceived creativity.

Out of the 545 toaster ideas, four were ideas for toasters that burn a holy image into the toast. Another four ideas were for toasters that optically detect burned toast. Figure 35 shows the ratings of clarity and creativity for each of these toaster idea sets as measured by a count of 2s. As the clarity score increase, the creativity score increases even though all the ideas in each set are essentially the same.

<center>94</center>
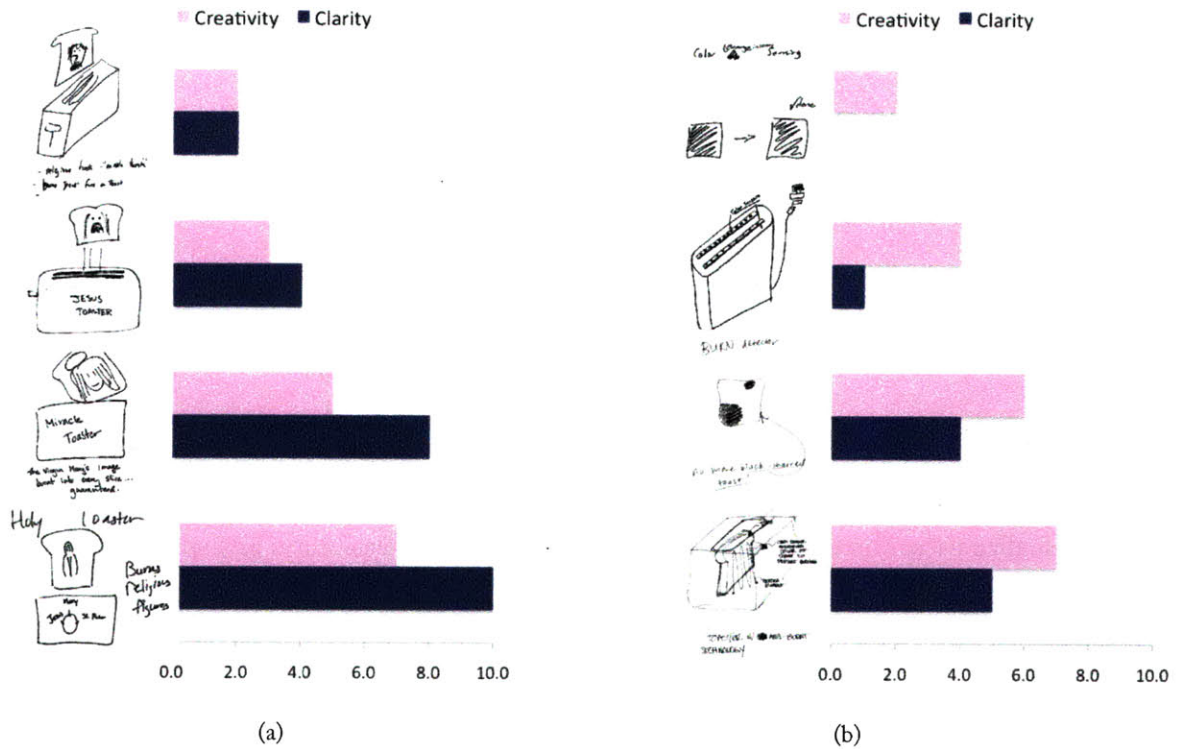
Figure 35. Creativity and Clarity of Holy Image Printing Toaster Ideas and Optical Burn Detecting Toaster Ideas

These findings suggest the importance of basic sketching ability for engineers. Creative ideas can be overlooked when they are poorly sketched.

5.4 A Map of Product Innovation

Innovation is the combination of knowledge or technologies in original and non-obvious valued new products, processes or services [31]. In essence, what makes a product idea innovative is a combination of novelty (or creativity), usefulness (or practical value), and feasibility. With strong correlations between creative and novelty scores (r^2=.80), we believe it is appropriate to rate *either* creativity or novelty as both would be redundant. This is inline with the novelty-based definitions of creativity and prior research [9, 23].
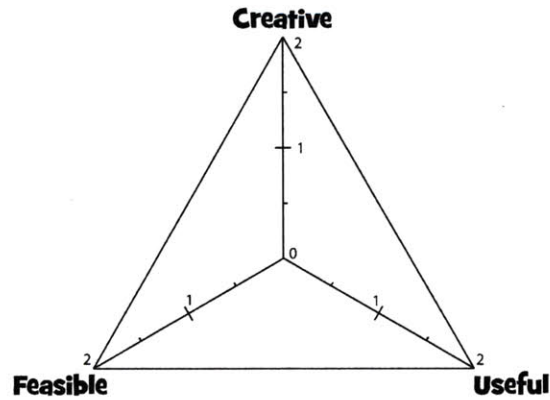
Figure 36. A Map of Product Innovation as a Spider Plot

We can visualize these three qualities (Creativity, Usefulness, Feasibility) as a spider plot as shown in Figure 36. Ideas can be mapped as a shape inside the triangle depending on their scores in each of these three areas. An idea that is truly innovative would have high scores in all three areas as shown in Figure 37. Note that an idea can be deemed creative without being of practical application, just as an idea can be deemed creative without being feasible.

(a)　　　　　　　　(b)
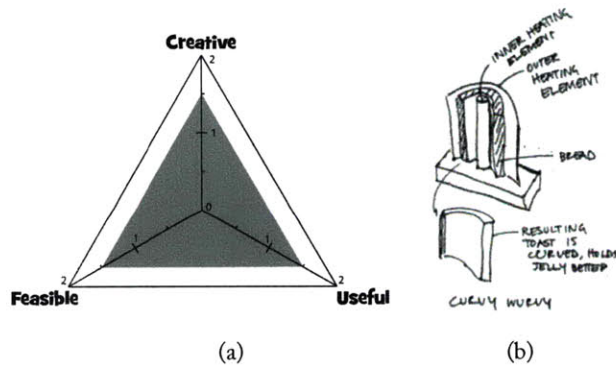
Figure 37. Example of an Innovative Product Idea Score and Innovative Idea Sketch

Ideas that are creative and feasible, but not useful can go by several names including: novelty items, aesthetic innovation [24], or chindogu [25]. These are products that you might find at Spencer's ®, The Sharper Image®, Think Geek® or Sky Mall ®. It could also include some art and items for play. These ideas do not have practical applications, but are valued and are feasible to be manufactured and marketed. These are depicted in Figure 38.
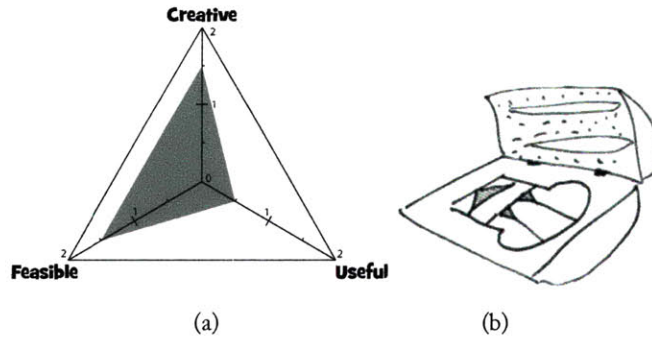
Figure 38. Example of a Novelty Item Idea Score and Novelty Item Idea Sketch

Novelty and usefulness are the two criteria that are most often associated with a "creative" idea [3, 47, 81]. Ideas that are creative and useful but not feasible are desired but less attainable. We call these ideas *chimera* or pipe dreams. These are depicted in Figure 39.
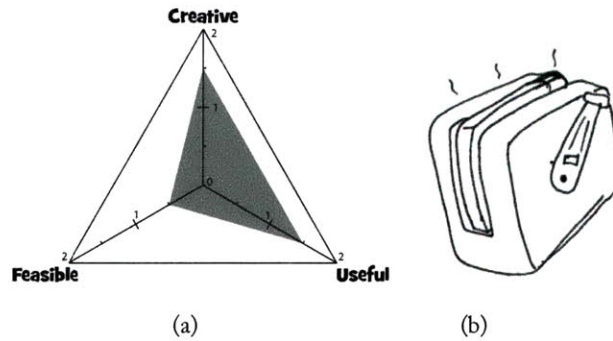
Figure 39. Example of a Chimera Idea Score and Chimera Idea Sketch

Ideas that are useful and feasible, but not novel are most likely existing products. This is the least helpful direction for a product ideation session, however it is a good staring point to build from. This is depicted in Figure 40.
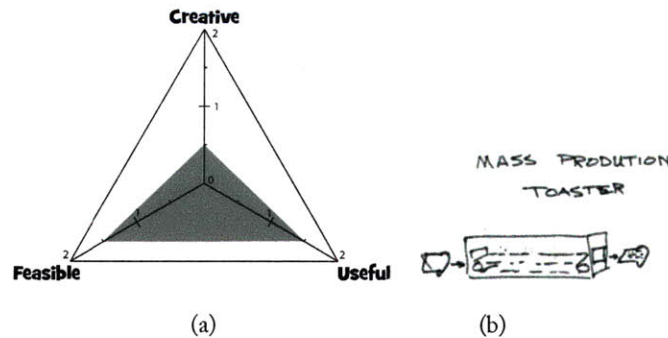
Figure 40. Example of an Existing Product Idea Score and Existing Product Idea Sketch

It is possible to have ideas that score high on only one of the three metrics. Ideas that are only novel/creative are *nonsense* ideas; these are non-obvious concepts that are not practical, useful or feasible. They might not even be related to the theme.

Creativity or novelty is of highest importance in blue-sky product design idea generation. Several studies, including this study, have found that novelty is the sole criteria for an idea to be considered creative [36, 37]. Usefulness and feasibility can be addressed after the idea generation session.

5.5 Reviewing Rate

Using Mechanical Turk, it takes between 2 and 3 minutes for each reviewer to rate 20 product ideas on 5 metrics. As it is online, multiple reviewers can rate product sets simultaneously. Scanning and naming the drawings is perhaps the most time intensive stage of the review, however, an automatic document scanner expedites the process. In future studies, only 3 metrics are required (creative or novel, useful, and feasible) to reasonably reduce hundreds of product concepts down to a more manageable set of promising ideas. Other studies have more elaborate sets of metrics [6, 12] some made of sub-metrics. These metrics are more appropriate for comparing a few product ideas in detail. To ask reviewers to rate hundreds of products on 9+ metrics would be time intensive and would cause fatigue [3].

5.6 Learning from Improvisational Comedy Training

Figure 23 implies that improvisational comedians are better than (or at least on par with) professional product designers when it comes to prolific "blue-sky" product idea generation. It could be that improvisational comedians are better at making many associations or they are less inhibited. Either way, perhaps designers can benefit from improvisational comedy training.

In this study, we had participants brainstorm nominally so we can target individual ideation ability rather than group interaction, which has more uncontrollable variables. We believe that group interaction skills can also be learned from improvisational comedy training, but this is out of the scope for this work.

We found that a group of 11 students individually generated on average 37% more ideas after an improvisational workshop. These students were highly fluent idea generators originally when compared to the test group of 26 MIT students from the prior study. Given the same time,

product theme, and prompt, the group of 11 students that signed up for the improvisation workshop/test had twice as many ideas on average *prior* to improvisation training. Perhaps the students that are willing to participate in an improvisation workshop are more open minded, willing to think differently and thus are going to be more prolific generators. After the workshop, these 11 students increased in idea fluency by approximately 37%. This shows that even highly fluent generators can become even more fluent with training.

5.7 Successful Caption Writing / Wording

The advice from New Yorker Cartoon Contest winners Larry Wood [116] and Patrick House [115] can be summed up in a few points:

Keep the captions simple and brief

Address what the subjects are thinking and feeling (Theory of Mind)

Incorporate everything in the cartoon

Do not use proper nouns and uncommon references

What Wood and House do not mention, which is perhaps most important, is resolving the incongruity in the cartoon in a non-obvious manner.

Figure 41 shows several captions from different subjects for the "Spacesuit in Office" cartoon. Each of these captions has the same basic message of "boss fires employee." These captions have very different average humor scores. We can use incongruity theory and the advice from Wood and House to explain the variance in scores.

The captions (D) "You're fired!" and (E) "*Boss*: You are fired; *Employee*: good thing I can hardly hear anything in this helmet" both do not resolve the incongruity of why the person is in a spacesuit in an office. Therefore, it seems appropriate that their humor scores were very low. The caption (C) "I'm sorry bubble-boy, you're fired" resolves the incongruity but does not resolve it in a non-obvious manner. This is similar to stating the obvious (*e.g. you are in a space suit and I'm firing you*). The captions (B) "I think we all saw this coming, you're being fired. Into Space" and (A) "You're fired in 3…2…1…" both resolve the incongruity in a non-obvious manner by playing off the dual meaning of with word "fired."

Simplicity also plays a role in the perceived humor. Captions A and B as well as captions D and E are very similar in nature. The most notable difference between these pairs is that captions A and D say that same thing in fewer words and thus they have higher scores than B and E. Less is more. Caption B could have been shorted to "you're being fired... into space" and would have had the same meaning with fewer words.

Linguistically, all of these captions have the same "connector" being "you're fired." However, only captions A and B have "disjunctors" that flip the meaning of the "connector." Without a "disjunctor" the joke lacks resolution [86].

Ways of Saying "You're Fired"

Caption	Average Score

You're fired in 3...2...1...

I think we all saw this coming, you're being fired. Into space.

I'm sorry bubble-boy, you're fired.

You're fired!

Boss: You are fired / Employee: ... (good thing I can hardly hear anythign in this helmet)
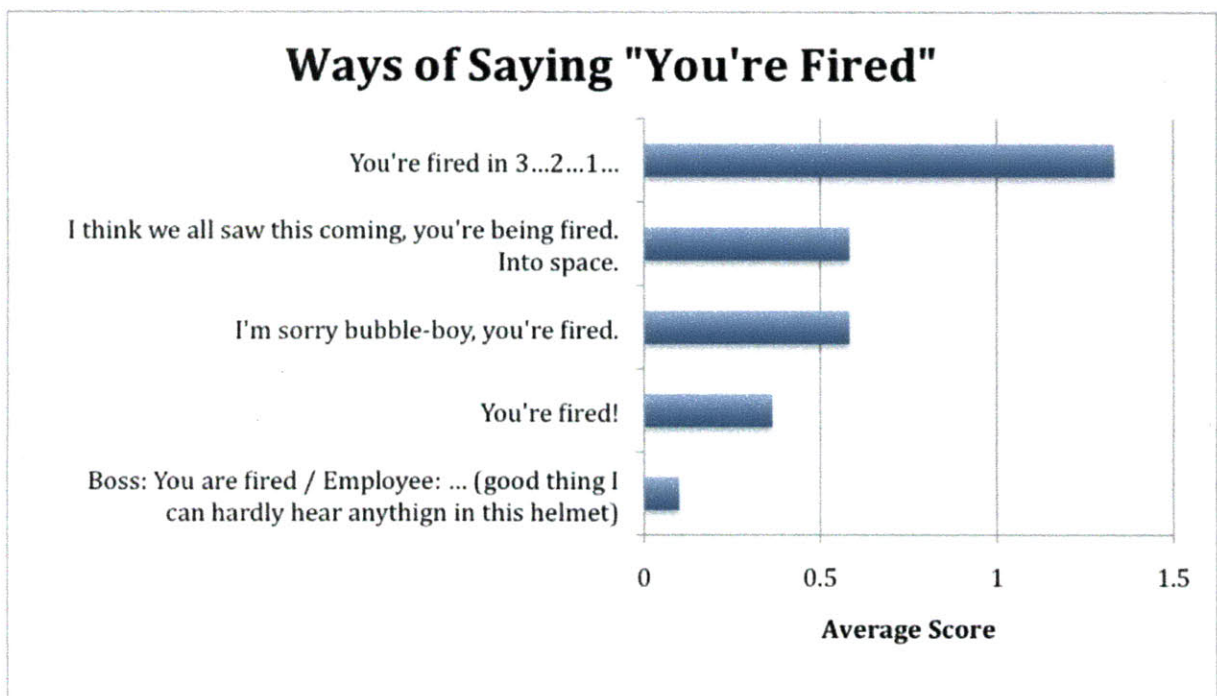
Average Score

Figure 41. Comparison of Scores for Similar Themed Captions

For the "Space Suit in Office" Cartoon, the average score was .69. (Surprisingly) the highest rated caption was "Yes, having beans for lunch was probably a bad idea" with a score of 1.6/2.0. The original published caption, "You know, Burkhart, if you're so damn afraid of the flu maybe you should just stay home," received a 1.0/2.0 average score. The distribution of average scores for all of the "Space Suit in Office" Cartoon captions is shown in Figure 42.
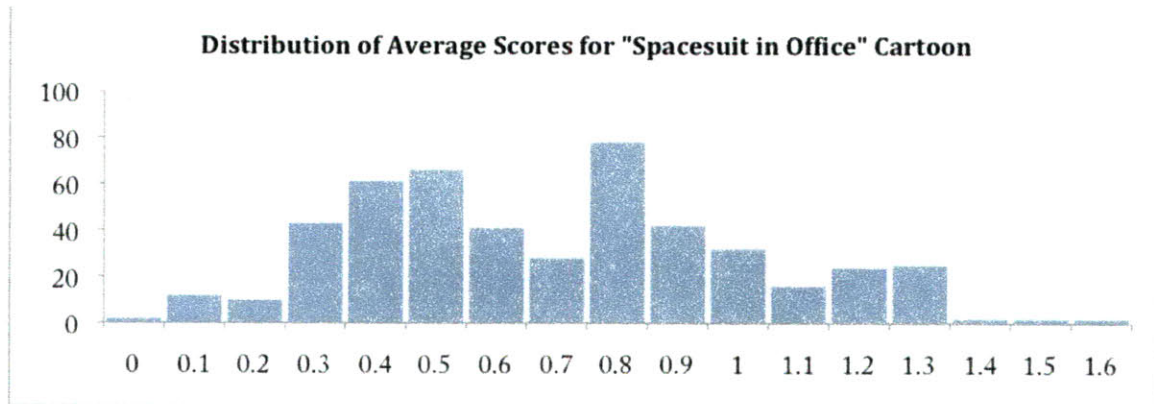
Figure 42. Distribution of Average Scores for "Spacesuit in Office" Cartoon

For the "Giant Toaster" Cartoon, the average score was 0.6. The highest rated caption was "The Incans were more sophisticated than we figured" with a score of 1.6/2.0. The original published caption, "At long last, Wyatt, our dream has come true and we are within reach of the legendary toast fields of the Sierra Madre," received a 0.5/2.0 average score. The distribution of average scores for all of the "Giant Toaster" Cartoon captions is shown in Figure 43.
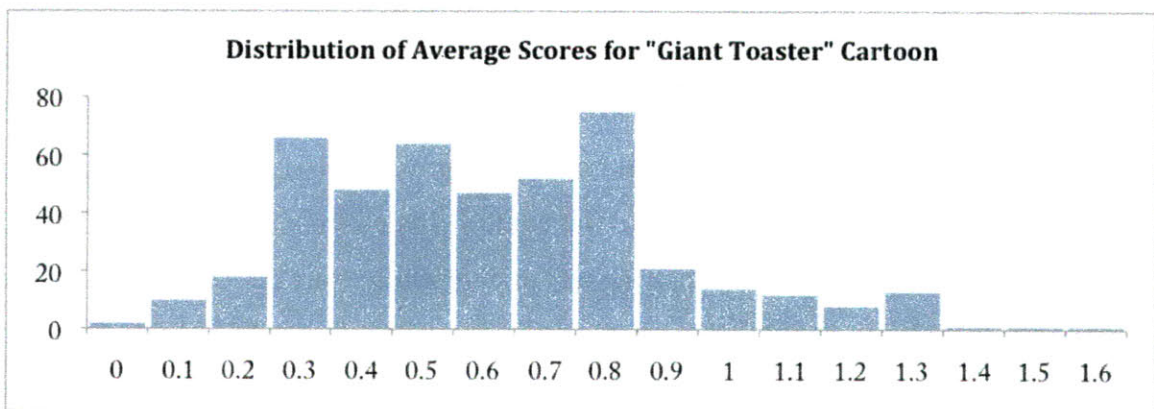
Figure 43. Distribution of Average Scores for "Giant Toaster" Cartoon

For the "Plant Psychologist" Cartoon, the average score was 0.61. The highest rated caption was "I think this all stems from my fear of insects" with a score of 1.7/2.0. The original published caption, "Do you think your failure to bloom could be caused not just by improper location but also by a fear of having your blooms compared with those of other African violets?" received a .75/2.0 average score. The distribution of average scores for all of the "Plant Psychologist" Cartoon captions is shown in Figure 44.

Distribution of Average Scores for "Plant Psychologist" Cartoon
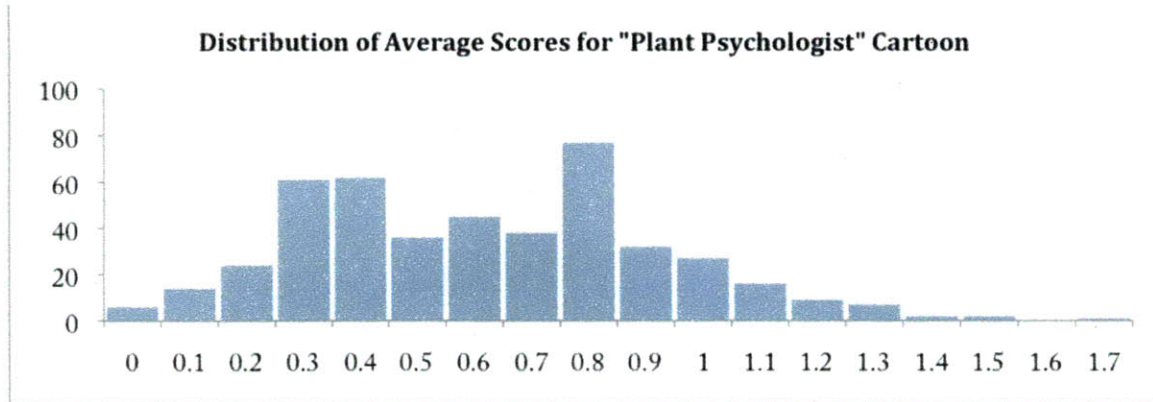
Figure 44. Distribution of Average Scores for "Plant Psychologist" Cartoon

The control captions being the original published cartoon captions scored quite low compared to many of the captions generated in this test. In the case of "Giant Toaster" caption, the original caption scored even lower than the average. These control captions were quite long and verbose and the essence of these jokes could have been condensed. Another theory is that the readers of the New Yorker have a different sense of humor than the individuals that are "workers" on Amazon Mechanical Turk. This was expressed by a winner of the New Yorker Cartoon Caption Competition, Patrick House, when he explains, "you are not trying to submit the funniest caption: you are trying to win The New Yorker's caption contest." One must understand the intended readership. "The reader [of the New Yorker] is isolated and introspective, probably on the train commuting to work…Laughing out loud is, in this context, an unseemly act… your caption should elicit, at best, a mild chuckle." Applying this to our results, a fart themed caption such as "Yes, having beans for lunch was probably a bad idea" is probably well suited for the majority of the internet surfing, Amazon Mechanical Turk workers, but is less suitable for the introspective New Yorker readership. In both humor and innovation, one must always keep the audience in mind.

There are things that are universally humorous and universally viewed as innovative, however, these fields are subjective to a greater degree than such fields as science and engineering.

6. Conclusions

6.1 Summary

Several studies have found correlations between humor and standardized creativity tests. Humor production, or wit, is perhaps the area that is most related to product design creativity. It is believed that both wit and product design innovation involves making non-obvious connections between seemingly unrelated things. In this study, we found that improvisational comedians on average produced 20% more product ideas and 25% more creative product ideas than professional product designers. Individual's quantity of product ideas produced was highly correlated (r^2=.82) with that individual's creativity scores and individual's quantity of cartoon captions produced was highly correlated (r^2=.64) with that individual's humor scores. It was also found that the few individuals that were highly fluent in prolific creative product ideation were also highly fluent in prolific humorous cartoon caption production. The majority of these prolific generators had improvisational comedy training. Many of the games used in introductory improvisation comedy classes are designed to train people in making prolific non-obvious connections between unrelated things. A workshop was developed composed of short-form association games to be administered prior to idea generation.

The work supports the notion that prolific idea generation *is* creative idea generation, as the quantity of ideas generated by the individual subjects had a very strong correlation with that subject's overall creativity scores (r^2=.82). This study also supports the novelty-based definition of creativity as reviewers' subjective ratings of idea creativity had a strong correlation with ratings of idea novelty (r^2=.80), but negligible correlation with idea usefulness (r^2=.16). We found that laymen reviewers were not able to accurately rate ideas for feasibility or marketability, and we suggest experts (engineers or designers) rate such categories. We found that sketch clarity affected the perceived creativity. This supports an argument that engineers should have basic drawing abilities so their creative ideas are not overlooked. Finally, we suggest three independent qualities that fully describe an innovative product idea: creative (as a subjective judgment), useful (as defined as having practical applications), and feasible (as determined by experts). We found that online reviewers are able to rate between 7-10 ideas per minute using these metrics. This is a rapid

means of taking a large collection of ideas and reducing them to a manageable set of the most promising ideas.

6.1 Future Work
Further testing is needed to better understand the impact of the improvisational training on the quality and quantity of ideas developed in brainstorming. Future work could explore what element of improvisational training is most important for idea generation (e.g. developing associative thinking skills, decreasing inhibitions or improving group dynamics). Other studies have found that positive affect is highly correlated with creativity. It is possible that the positive energy generated from improvisational warm-up plays a larger role in increasing idea fluency than the other elements of improvisational training. In this same direction, empirical studies are needed to explore the affect of improvisational training on group brainstorming as opposed to individual idea generation.

Although improvisational comedians did produce highly creative product ideas, it could be argued that product design idea generation is more of a visual-spatial creativity whereas cartoon captions and improvisational humor involve more of a verbal-linguistic creativity. Perhaps there are two independent styles for making associations those being visualizing and verbalizing where the former involves concepts and the later involves language [3]. This would be in line with the Theory of Multiple Intelligences in which an individual can be intelligent or creative in one domain and not another [148]. It has been argued that there is no empirical research supporting the Theory of Multiple Intelligences [149]. Future work could investigate what "types of intelligence" are associated with idea generation and humor production.

A similar argument could be made that the cartoon caption test is word based, while the idea generation test is sketching based. The product ideation test involves communication through drawing, while the caption test involves communication through words. These two tests are almost opposites in nature: the caption test involves explaining a drawing with a concept while the product ideation involves explaining a concept with a drawing. It is possible that participants that are good at presenting ideas with drawings are not good at presenting information with words. Turner suggests that a distinction be made between the generation of humor (i.e. finding an incongruity or novel connection) and the effective communication of the humor (i.e. the wording) [110]. A prior study found that the cartoon caption test (verbal humor test) and Torrance's Circles

104

(visual creativity test) were not significantly correlated (r = +.08) [11]. It may be interesting to perform this correlation study with a set of professional artists and a set of professional writers to see how drawing and writing abilities influence the humor and creativity of the captions and product ideas. A future study could compare written ideas to written captions; however, we believe more words would be needed to accurately describe a concept.

References

1.   L'Engle, M., *A wrinkle in time*. A Yearling book. 1962, New York: Dell Pub. Co. [10], 211, [2] p.

2.   Nussbaum, B., R. Berner, and D. Brady, *Get Creative! How to Build Innovative Companies*, in *BuisnessWeek*. 2005, Bloomberg L. P.

3.   Mednick, S.A., *The associative basis of the creative process*. Psychol Rev, 1962. **69**: p. 220-32.

4.   Martindale, C., *Biological Basis of Creativity*, in *Handbook of Creativity*, R.J. Sternberg, Editor. 1999, Cambridge University Press: Cambridge. p. 137-152.

5.   Koestler, A., *The act of creation*. 1964, New York,: Macmillan. 751 p.

6.   Goodchilds, J.D., *On Being Witty: Causes, Correlates and Consequences*, in *The Psychology of Humor*, J.H.G.a.P.E. McGhee, Editor. 1972, Academic Press: New York. p. 173-193.

7.   Besemer, S.P. and D.J. Treffinger, *Analysis of Creative Products - Review and Synthesis*. Journal of Creative Behavior, 1981. **15**(3): p. 158-178.

8.   O'Quin, K. and P. Derks, *Humor*, in *Encyclopedia of Creativity*, S.R. Pritzker and M.A. Runco, Editors. 1999, Academic Press: San Diego. p. 845-853.

9.   Smith, E.E. and H.L. White, *WIT, CREATIVITY, AND SARCASM*. Journal of Applied Psychology, 1965. **49**(2): p. 131-134.

10.   Treadwell.Y, *HUMOR AND CREATIVITY*. Psychological Reports, 1970. **26**(1): p. 55-&.

11.   Babad, E.Y., *MULTI-METHOD APPROACH TO ASSESSMENT OF HUMOR - CRITICAL LOOK AT HUMOR TESTS*. Journal of Personality, 1974. **42**(4): p. 618-631.

12.   Rouff, L.L., *CREATIVITY AND SENSE OF HUMOR*. Psychological Reports, 1975. **37**(3): p. 1022-1022.

13.   Brodzinsky, D.M. and J. Rubien, *HUMOR PRODUCTION AS A FUNCTION OF SEX OF SUBJECT, CREATIVITY, AND CARTOON CONTENT*. Journal of Consulting and Clinical Psychology, 1976. **44**(4): p. 597-600.

14.   Clabby, J.F., *THE WIT - A PERSONALITY ANALYSIS*. Journal of Personality Assessment, 1980. **44**(3): p. 307-310.

15.   Humke, C. and C.E. Schaefer, *Sense of humor and creativity*. Perceptual and Motor Skills, 1996. **82**(2): p. 544-546.

16.   Danzico, L. *Frames: Notes on Interaction and Design*. in *Interaction10*. 2010. Savannah, GA.

17.   Gerber, E., *Using Improvisation to Enhance the Effectiveness of Brainstorming*. Chi2009: Proceedings of the 27th Annual Chi Conference on Human Factors in Computing Systems, Vols 1-4, 2009: p. 97-104.

18.   Nicolson, S.H., *The English Sense of Humour, An Essay*. 1946, London: The Dropmore Press.

19.   Simpson, J.A. and E.S.C. Weiner, *The Oxford English dictionary*. 2nd ed. 1989, Oxford
New York: Clarendon Press ;
Oxford University Press.

20.   Martin, R.A., *The psychology of humor : an integrative approach*. 2007, Amsterdam ; Boston: Elsevier Academic Press. xviii, 446 p.

21.   Feingold, A. and R. Mazzella, *PRELIMINARY VALIDATION OF A MULTIDIMENSIONAL MODEL OF WITTINESS*. Journal of Personality, 1993. 61(3): p. 439-456.

22.   Fink, A. and A.C. Neubauer. *Neuroscientific Approaches to the Study of Creativity*. in *International Workshop on Studying Design Creativity*. 2008.

23.   Keith-Spiegel, P., *Early Conceptions of Humor: Varieties and Issues*, in *The Psychology of Humor*, J.H. Goldstein and P.E. McGhee, Editors. 1972, Academic Press: New York. p. 4-39.

24.   Suls, J.M., *A Two-Stage Model for the Appreciation of Jokes and Cartoons: An Information-Processing Analysis*, in *The Psychology of Humor*, J.H. Goldstein and P.E. McGhee, Editors. 1972, Academic Press: New York. p. 81-100.

25.   Sitton, S.C. and E.R. Pierce, *Synesthesia, creativity and puns*. Psychological Reports, 2004. 95(2): p. 577-580.

26.   Sternberg, R.J. and T.I. Lubart, *The Concept of Creativity: Prospects and Paradigms*, in *Handbook of Creativitity*, R.J. Sternberg, Editor. 1999, Cambridge University Press: New York.

27.   Mayer, R.E., *Fifty Years of Creativity Research*, in *Handbook of Creativity*, R.J. Sternberg, Editor. 1999, Cambridge University Press: London. p. 449-460.

28.   Amabile, T., *Creativity in context*. 1996, Boulder, Colo.: Westview Press. xviii, 317 p.

29.   Luecke, R. and R. Katz, *Managing Creativity and Innovation*. 2003, Boston: Harvard Business School Press.

30.   Garcia, R. and R. Calantone, *A critical look at technological innovation typology and innovativeness terminology: a literature review*. Journal of Product Innovation Management, 2002. 19(2): p. 110-132.

31.   Ashby, M.F. and K. Johnson, *Materials and design : the art and science of material selection in product design*. 2002, Oxford ; Boston: Butterworth-Heinemann. vii, 336 p.

32.   Shah, J.J., S.V. Kulkarni, and N. Vargas-Hernandez, *Evaluation of idea generation methods for conceptual design: Effectiveness metrics and design of experiments.* Journal of Mechanical Design, 2000. **122**(4): p. 377-384.

33.   Yu, F.T., *Novelty and Its Acceptance: An Intersubjective Perspective.* International Journal of Arts and Technology, 2008. **1**(1): p. 119-130.

34.   Hekkert, P., D. Snelders, and P.C.W. van Wieringen, *'Most advanced, yet acceptable': Typicality and novelty as joint predictors of aesthetic preference in industrial design.* British Journal of Psychology, 2003. **94**: p. 111-124.

35.   Saunders, R. and J.S. Gero, *Designing for interest and novelty - Motivating design agents.* Computer Aided Architectural Design Futures 2001, Proceedings, 2001: p. 725-738.

36.   Runco, M.A. and R.E. Charles, *Judgments of Originality and Appropriateness as Predictors of Creativity.* Personality and Individual Differences, 1993. **15**(5): p. 537-546.

37.   Christiaans, H., *Creativity as a design criterion.* Creativity Research Journal, 2002. **14**(1): p. 41-54.

38.   Eisenman, M.L., *Essays on Aesthetic Innovation*, in *Graduate School of Business.* 2006, Columbia University: NY.

39.   Norman, D.A., *Emotional design : why we love (or hate) everyday things.* 2004, New York: Basic Books. x, 257 p.

40.   Kudrowitz, B.M. and D.R. Wallace, *The Play Pyramid: A Play Classification and Ideation Tool for Toy Design.* International Journal of Arts and Technology, 2010. **3**(1): p. 36-56.

41.   Patton, A. and R. Bannerot. *Chindogu: A Problem Solving Strategy for Transforming Uselessness into Fearlessness.* in *Annual Conference of the American Soceity for Engineering Education Gulf Southwest Section.* 2002. Lafayette.

42.   Kawakami, K., *THe Big Bento Box of Unuseless Japanese Inventions.* 1995, New York: Norton.

43.   Ulrich, K.T. and S.D. Eppinger, *Product design and development.* 4th ed. 2008, Boston: McGraw-Hill Higher Education. xv, 368 p.

44.   Niku, S., *Creative Product Design.* 2009, Hobeken, NJ: John WIley and Sons, Inc.

45.   Geisel, T., *Oh, The Things You Can Think!* 1975, New York: Random House.

46.   Smith, G.F., *Idea-generation techniques: A formulary of active ingredient.* Journal of Creative Behavior, 1998. **32**(2): p. 107-133.

47.   Butler, D.L. and M.A. Kline, *Good versus creative solutions: A comparison of brainstorming, hierarchical, and perspective-changing heuristics.* Creativity Research Journal, 1998. **11**(4): p. 325-331.

48.   Michalko, M., *Thinkertoys : a handbook of creative-thinking techniques*. 2nd ed. 2006, Berkeley, Calif.: Ten Speed Press. xx, 394 p.

49.   Von Oech, R., *A Whack on the Side of the Head*. 1983, Menlo Park: Creative Think.

50.   De Bono, E., *Lateral thinking: a textbook of creativity*. 1970, London,: Ward Lock Educational. 300 p.

51.   De Bono, E., *Serious creativity : using the power of lateral thinking to create new ideas*. 1st ed. 1992, New York, N.Y.: HarperBusiness. xiv, 338 p.

52.   De Bono, E., *Six thinking hats*. 1st Back Bay pbk. ed. 1999, Boston: Back Bay Books. xiii, 173 p.

53.   Lieberman, J.N., *Playfulness: Its Relationship to Imagination and Creativity*. 1977: Academic Press Inc.

54.   Dansky, J.L. and Silverma.Iw, *EFFECTS OF PLAY ON ASSOCIATIVE FLUENCY IN PRESCHOOL-AGED CHILDREN*. Developmental Psychology, 1973. 9(1): p. 38-43.

55.   Dansky, J., *Play*, in *Encyclopedia of Creativity*, S.R. Pritzker and M. Runco, Editors. 1999, Academic Press: San Diego.

56.   Berretta, S. and G. Privette, *INFLUENCE OF PLAY ON CREATIVE-THINKING*. Perceptual and Motor Skills, 1990. 71(2): p. 659-666.

57.   Isen, A.M., K.A. Daubman, and G.P. Nowicki, *POSITIVE AFFECT FACILITATES CREATIVE PROBLEM-SOLVING*. Journal of Personality and Social Psychology, 1987. 52(6): p. 1122-1131.

58.   Isen, A.M., et al., *THE INFLUENCE OF POSITIVE AFFECT ON THE UNUSUALNESS OF WORD-ASSOCIATIONS*. Journal of Personality and Social Psychology, 1985. 48(6): p. 1413-1426.

59.   Russ, S., *Affect and Creativity: the Role of Affect and Play in the Creative Process*. 1993, Hillsdale, NJ: Lawrence Erlbaum Associates.

60.   Maccrimmon, K.R. and C. Wagner, *Stimulating Ideas Through Creativity Software*. Management Science, 1994. 40(11): p. 1514-1532.

61.   Ludden, G.D.S., et al., *Surprise & humor in product design. Designing sensory metaphors in multiple modalities*. Humor, in press.

62.   Buzan, T. and B. Buzan, *The mind map book : how to use radiant thinking to maximize your brain's untapped potential*. 1993, New York: Plume. 320 p.

63.   Osborn, A.F., *Applied imagination; principles and procedures of creative problem-solving*. 3d rev. ed. 1963, New York,: Scribner. 417 p.

64.   Diehl, M. and W. Stroebe, *Productivity Loss in Brainstorming Groups - Toward the Solution of a Riddle*. Journal of Personality and Social Psychology, 1987. 53(3): p. 497-509.

65.   Bray, R.M., N.L. Kerr, and R.S. Atkin, *Effects of Group-Size, Problem Difficulty, and Sex on Group-Performance and Member Reactions.* Journal of Personality and Social Psychology, 1978. 36(11): p. 1224-1240.

66.   Rietzschel, E.F., B.A. Nijstad, and W. Stroebe, *Productivity is not enough: A comparison of interactive and nominal brainstorming groups on idea generation and selection.* Journal of Experimental Social Psychology, 2006. 42(2): p. 244-251.

67.   Silverstein, S., *Falling up : poems and drawings.* 1996, New York, N.Y.: HarperCollins. 171 p.

68.   Amabile, T.M., *Social-Psychology of Creativity - A Consensual Assessment Technique.* Journal of Personality and Social Psychology, 1982. 43(5): p. 997-1013.

69.   Torrance, E.P., *Predictive Validity of Torrance Tests of Creative Thinking.* Journal of Creative Behavior, 1972. 6(4): p. 236-252.

70.   Guilford, J.P., *The Structure of Intellect.* Psychological Bulletin, 1956. 53(4): p. 267-293.

71.   Forbach, G.B. and R.G. Evans, *THE REMOTE ASSOCIATES TEST AS A PREDICTOR OF PRODUCTIVITY IN BRAINSTORMING GROUPS.* Applied Psychological Measurement, 1981. 5(3): p. 333-339.

72.   Taft, R. and J.R. Rossiter, *REMOTE ASSOCIATES TEST - DIVERGENT OR CONVERGENT THINKING.* Psychological Reports, 1966. 19(3P2): p. 1313-&.

73.   O'Quin, K. and S. Besemer, *Creative Products,* in *Encyclopedia of Creativity,* S.R. Pritzker and M. Runco, Editors. 1999, Academic Press: San Diego. p. 267-278.

74.   Dean, D.L., et al., *Identifying quality, novel, and creative ideas: Constructs and scales for idea evaluation.* Journal of the Association for Information Systems, 2006. 7(10): p. 646-698.

75.   Besemer, S.P., *Creative product analysis matrix: Testing the model structure and a comparison among products - Three novel chairs.* Creativity Research Journal, 1998. 11(4): p. 333-346.

76.   Besemer, S. and K. Oquin, *Analyzing Creative Products - Refinement and Test of a Judging Instrument.* Journal of Creative Behavior, 1986. 20(2): p. 115-126.

77.   Besemer, S.P. and K. O'Quin, *Confirming the three-factor creative product analysis matrix model in an American sample.* Creativity Research Journal, 1999. 12(4): p. 287-296.

78.   Horn, D. and G. Salvendy, *Consumer-based assessment of product creativity: A review and reappraisal.* Human Factors and Ergonomics in Manufacturing, 2006. 16(2): p. 155-175.

79.   Horn, D. and G. Salvendy, *Measuring Consumer Perception of Product Creativity: Impact on Satisfaction and Purchasability.* Human Factors and Ergonomics in Manufacturing, 2009. 19(3): p. 223-240.

80.   Ward, W.C. and P.W. Cox, *Field Study of Nonverbal Creativity.* Journal of Personality, 1974. 42(2): p. 202-219.

81.   Shah, J.J. and N. Vargas-Hernandez, *Metrics for Measuring Ideation Effectiveness.* Design Studies, 2003. **24:** p. 111-134.

82.   Reinig, B.A., R.O. Briggs, and J.F. Nunamaker, *On the measurement of ideation quality.* Journal of Management Information Systems, 2007. **23**(4): p. 143-161.

83.   Briggs, R.O., et al., *Quality as a function of quantity in electronic brainstorming.* Thirtieth Hawaii International Conference on System Sciences, Vol 2, 1997: p. 94-103.

84.   Reinig, B.A. and R.O. Briggs. *Measuring the Quality of Ideation Technology and Techniques.* in *HICSS-39.* 2006. Kauai, Hawaii.

85.   White, E.B. and K.S.A. White, *A subtreasury of American humor.* 1941, New York,: Coward-McCann. xxxii p., 2 l., 3-814 p.

86.   Attardo, S., *Linguistic theories of humor.* Humor research 1. 1994, Berlin ; New York: Mouton de Gruyter. xix, 426 p.

87.   Cundall, M.K., *Humor and the limits of incongruity.* Creativity Research Journal, 2007. **19**(2-3): p. 203-211.

88.   Freud, S. and J. Strachey, *Jokes and their relation to the unconscious.* The Norton library. 1963, New York: W. W. Norton. 258 p.

89.   Shultz, T.R., *ROLE OF INCONGRUITY AND RESOLUTION IN CHILDRENS APPRECIATION OF CARTOON HUMOR.* Journal of Experimental Child Psychology, 1972. **13**(3): p. 456-&.

90.   Kenny, D.T., *The contingency of humor appreciation on the stimulus-confirmation of joke-ending expectations.* Jour Abnormal and Social Psychol, 1955. **51-31**((2)): p. 644-648.

91.   McGhee, P.E., *COGNITIVE DEVELOPMENT AND CHILDRENS COMPREHENSION OF HUMOR.* Child Development, 1971. **42**(1): p. 123-&.

92.   Derks, P.L., H.M. Leichtman, and P.J. Carroll, *PRODUCTION AND JUDGMENT OF HUMOR BY SCHIZOPHRENICS AND COLLEGE-STUDENTS.* Bulletin of the Psychonomic Society, 1975. **6**(3): p. 300-302.

93.   Martin, R.A., *Approaches to the Sense of Humor,* in *The Sense of Humor: Explorations of a Personality Characterisitc,* W. Ruch, Editor. 1998, Mouton de Gruyter: Berlin.

94.   Sawyer, R.K., *Improvisation,* in *Encyclopedia of Creativity,* S.R. Pritzker and M. Runco, Editors. 1999, Academic Press: San Diego. p. 845-853.

95.   Salinsky, T. and D. Frances-White, *The Improv Handbook.* 2008, New York: Continuum.

96.   Spolin, V., *Improvisation for the theater : a handbook of teaching and directing techniques.* 3rd ed. 1999, Evanston, Ill.: Northwestern University Press. lv, 412 p.

97.   Fry, W.F., *POWER OF POLITICAL HUMOR.* Journal of Popular Culture, 1976. **10**(1): p. 227-231.

98.   Oconnell, W.E., *CREATIVITY IN HUMOR.* Journal of Social Psychology, 1969. **78**(2): p. 237-&.

99.   Oconnell, W.E., *THE ADAPTIVE FUNCTIONS OF WIT AND HUMOR.* Journal of Abnormal and Social Psychology, 1960. **61**(2): p. 263-270.

100.   Ruch, W., *Measurement approaches to the sense of humor: Introduction and overview.* Humor-International Journal of Humor Research, 1996. **9**(3-4): p. 239-250.

101.   Kohler, G. and W. Ruch, *Sources of variance in current sense of humor inventories: How much substance, how much method variance?* Humor-International Journal of Humor Research, 1996. **9**(3-4): p. 363-397.

102.   Ruch, W., *The sense of humor : explorations of a personality characteristic.* Humor research 3. 1998, Berlin ; New York: Mouton de Gruyter. x, 498 p.

103.   Svebak, S., *Revised questionnaire on the sense of humor.* Scand J Psychol, 1974. **15**(4): p. 328-31.

104.   Martin, R.A. and H.M. Lefcourt, *Sense of Humor as a Moderator of the Relation between Stressors and Moods.* Journal of Personality and Social Psychology, 1983. **45**(6): p. 1313-1324.

105.   Martin, R.A. and H.M. Lefcourt, *Situational Humor Response Questionnaire - Quantitative Measure of Sense of Humor.* Journal of Personality and Social Psychology, 1984. **47**(1): p. 145-155.

106.   Ruch, W. and F.J. Hehl, *INTOLERANCE OF AMBIGUITY AS A FACTOR IN THE APPRECIATION OF HUMOR.* Personality and Individual Differences, 1983. **4**(5): p. 443-449.

107.   Thorson, J.A. and F.C. Powell, *Development and Validation of a Multidimensional Sense of Humor Scale.* Journal of Clinical Psychology, 1993. **49**(1): p. 13-23.

108.   Craik, K.H., M.D. Lampert, and A.J. Nelson, *Sense of humor and styles of everyday humorous conduct.* Humor-International Journal of Humor Research, 1996. **9**(3-4): p. 273-302.

109.   Martin, R.A., et al., *Individual differences in uses of humor and their relation to psychological well-being: Development of the Humor Styles Questionnaire.* Journal of Research in Personality, 2003. **37**(1): p. 48-75.

110.   Turner, R.G., *SELF-MONITORING AND HUMOR PRODUCTION.* Journal of Personality, 1980. **48**(2): p. 163-172.

111.   Ziller, R.C., R.D. Behringer, and J.D. Goodchilds, *GROUP CREATIVITY UNDER CONDITIONS OF SUCCESS OR FAILURE AND VARIATIONS IN GROUP STABILITY.* Journal of Applied Psychology, 1962. **46**(1): p. 43-49.

112.   Derks, P. and D. Hervas, *CREATIVITY IN HUMOR PRODUCTION - QUANTITY AND QUALITY IN DIVERGENT THINKING.* Bulletin of the Psychonomic Society, 1988. **26**(1): p. 37-39.

113. Derks, P., *CATEGORY AND RATIO SCALING OF SEXUAL AND INNOCENT CARTOONS*. Humor-International Journal of Humor Research, 1992. 5(4): p. 319-329.

114. Derks, P.L., *HUMOR PRODUCTION - AN EXAMINATION OF 3 MODELS OF CREATIVITY*. Journal of Creative Behavior, 1987. 21(4): p. 325-326.

115. House, P., *How to Win the New Yorker Cartoon Caption Competition*, in *Slate*. 2008, Washington Post.Newsweek Interactive Co. LLC.

116. Altman, A., *How to Win the New Yorker Caption Contest*, in *The New Yorker*. 2009: New york.

117. McCloud, S., *Understanding comics : the invisible art*. 1st HarperPerennial ed. 1994, New York: HarperPerennial. 215 , [1] p.

118. McCloud, S., *Reinventing comics*. 1st Perennial ed. 2000, New York: Perennial. 237 p.

119. McCloud, S., *Making Comics*. 2006, New York: Harper Paperbacks.

120. Bartolo, A., et al., *Humor comprehension and appreciation: An fMRI study*. Journal of Cognitive Neuroscience, 2006. 18(11): p. 1789-1798.

121. Taber, K.H., M. Redden, and R.A. Hurley, *Functional anatomy of humor: positive affect and chronic mental illness*. J Neuropsychiatry Clin Neurosci, 2007. 19(4): p. 358-62.

122. Flaherty, A.W., *Frontotemporal and dopaminergic control of idea generation and creative drive*. Journal of Comparative Neurology, 2005. 493(1): p. 147-153.

123. Block, R.I. and J.R. Wittenborn, *MARIJUANA EFFECTS ON ASSOCIATIVE PROCESSES*. Psychopharmacology, 1985. 85(4): p. 426-430.

124. Gur, R.C. and J. Reyher, *ENHANCEMENT OF CREATIVITY VIA FREE-IMAGERY AND HYPNOSIS*. American Journal of Clinical Hypnosis, 1976. 18(4): p. 237-249.

125. Weckowicz, T.E., et al., *Effect of marijuana on divergent and convergent production cognitive tests*. J Abnorm Psychol, 1975. 84(4): p. 386-98.

126. Hines, D. and C. Martindale, *Induced lateral eye-movements and creative and intellectual performance*. Percept Mot Skills, 1974. 39(1): p. 153-4.

127. Bowden, E.M. and M. Jung-Beeman, *Getting the Right Idea: Semantic Activiation in the Right Hemisphere May Help Solve Insight Problems*. Psychological Science, 1998. 9(6): p. 435-440.

128. Johnson, A.M., *A study of humor and the right hemisphere*. Percept Mot Skills, 1990. 70(3 Pt 1): p. 995-1002.

129.	Shammi, P. and D.T. Stuss, *Humour appreciation: a role of the right frontal lobe.* Brain, 1999. **122**: p. 657-666.

130.	Brownell, H.H., et al., *Surprise but not coherence: sensitivity to verbal humor in right-hemisphere patients.* Brain Lang, 1983. **18**(1): p. 20-7.

131.	Bihrle, A.M., et al., *Comprehension of humorous and nonhumorous materials by left and right brain-damaged patients.* Brain Cogn, 1986. **5**(4): p. 399-411.

132.	Howard-Jones, P.A., et al., *Semantic divergence and creative story generation: an fMRI investigation.* Brain Res Cogn Brain Res, 2005. **25**(1): p. 240-50.

133.	Seger, C.A., et al., *Functional magnetic resonance imaging evidence for right-hemisphere involvement in processing unusual semantic relationships.* Neuropsychology, 2000. **14**(3): p. 361-369.

134.	Snyder, A., T. Bossomaier, and D.J. Mitchell, *Concept formation: 'Object' attributes dynamically inhibited from conscious awareness.* Journal of Integrative Neuroscience, 2004. **3**(1): p. 31-46.

135.	Mashal, N., et al., *An fMRI investigation of the neural correlates underlying the processing of novel metaphoric expressions.* Brain Lang, 2007. **100**(2): p. 115-26.

136.	Morris, W., *The American Heritage dictionary of the English language.* New college ed. 1976, Boston: Houghton Mifflin. L, 1550 p.

137.	Hocking, J. and C.J. Price, *The role of the posterior superior temporal sulcus in audiovisual processing.* Cereb Cortex, 2008. **18**(10): p. 2439-49.

138.	Bowden, E.M. and M. Jung-Beeman, *Aha! - Insight experience correlates with solution activation in the right hemisphere.* Psychonomic Bulletin & Review, 2003. **10**(3): p. 730-737.

139.	Limb, C.J. and A.R. Braun, *Neural Substrates of Spontaneous Musical Performance: An fMRI Study of Jazz Improvisation.* Plos One, 2008. **3**(2): p. 9.

140.	Berkowitz, A.L. and D. Ansari, *Expertise-related deactivation of the right temporoparietal junction during musical improvisation.* Neuroimage, 2010. **49**(1): p. 712-719.

141.	Fink, A., B. Graif, and A.C. Neubauer, *Brain correlates underlying creative thinking: EEG alpha activity in professional vs. novice dancers.* Neuroimage, 2009. **46**(3): p. 854-862.

142.	Amabile, T., *The social psychology of creativity.* Springer series in social psychology. 1983, New York: Springer-Verlag. xiii, 245 p.

143.	Bowden, E.M. and M. Jung-Beeman, *Normative data for 144 compound remote associate problems.* Behavior Research Methods Instruments & Computers, 2003. **35**(4): p. 634-639.

144.   Ziv, A., *The Influence of Humorous Atmosphere on Divergent Thinking.* Contemporary Educational Psychology, 1983. **8**: p. 68-75.

145.   Ziv, A., *Facilitating effects of humor on creativity.* J Educ Psychol, 1976. **68**(3): p. 318-22.

146.   Dugosh, K.L., et al., *Cognitive stimulation in brainstorming.* J Pers Soc Psychol, 2000. **79**(5): p. 722-35.

147.   ImprovEncyclopedia. *Improv Games.* 2007 [cited 2010 March 5]; Available from: http://improvencyclopedia.org/games.

148.   Gardner, H., *Frames of mind : the theory of multiple intelligences.* 10th anniversary ed. 1993, New York, NY: BasicBooks. xxxii, 440 p.

149.   Waterhouse, L., *Inadequate evidence for multiple intelligences, Mozart effect, and emotional intelligence theories.* Educational Psychologist, 2006. **41**(4): p. 247-255.

Appendix A

30 Question Rat Test with Solutions and Number of Participants (out of 100) with a Correct Response adapted from Bowden and Beeman, 2003

#	Words	Solution	Number
1	night wrist stop	watch	82
2	piece mind dating	game	53
3	peach arm tar	pit	41
4	flower friend scout	girl	51
5	dew comb bee	honey	80
6	food forward break	fast	41
7	basket eight snow	ball	39
8	break bean cake	coffee	12
9	worm shelf end	book	53
10	shine beam struck	moon	41
11	duck fold dollar	bill	80
12	pie luck belly	pot	49
13	down question check	mark	28
14	roll bean fish	jelly	11
15	grass king meat	crab	13
16	fountain baking pop	soda	78
17	rocking wheel high	chair	80
18	catcher food hot	dog	30
19	print berry bird	blue	49
20	wise work tower	clock	13
21	test runner map	road	18
22	health taker less	care	29
23	man glue star	super	18
24	dive light rocket	sky	18
25	right cat carbon	copy	39
26	dress dial flower	sun	29
27	force line mail	air	28
28	rain test stomach	acid	22
29	fence card master	post	18
30	cry front ship	battle	11

Appendix B

Survey Page 1

Wit and Innovation Survey 2008

you may elect not to answer any question

ID Number _____ Age _____ Gender: M F

Circle the Most Appropriate:

Student (Freshman)   Student (Senior)   Design Industry   Improv Comedy   Other _____

Years of experience (in design industry or with improv comedy) _____

Have you participated in a planned idea generation or brainstorming session before? Y / N

In which of these tests did you feel more creative? Cartoon Caption   Product Ideation   Neither

Which test did you take first?   Cartoon Caption   Product Ideation

Did you view the caption test as a game? Y / N

Did you view the ideation as a game? Y / N

I practice the arts (visual, written, musical, performance, etc):
     never   rarely   at least once/month   at least once/week   daily

I do something creative:
     never   rarely   at least once/month   at least once/week   daily

I make others laugh:
     never   rarely   at least once/month   at least once/week   daily

I build things:
     never   rarely   at least once/month   at least once/week   daily

I am generally a creative person
     strongly disagree   disagree   neutral/no opinion   agree   agree strongly

I am generally a witty person
     strongly disagree   disagree   neutral/no opinion   agree   agree strongly

Overall I am satisfied with the cartoon captions I created
     strongly disagree   disagree   neutral/no opinion   agree   agree strongly

Overall I am satisfied with the product ideas I created
     strongly disagree   disagree   neutral/no opinion   agree   agree strongly

This experience was enjoyable
     strongly disagree   disagree   neutral/no opinion   agree   agree strongly

This experience was stressful
     strongly disagree   disagree   neutral/no opinion   agree   agree strongly

Where would you rate the caption test on a scale from Work to Play?
     work   work disguised as play   directed play   guided play   free play

Where would you rate the brainstorming on a scale from Work to Play?
     work   work disguised as play   directed play   guided play   free play

Appendix C
Survey Page 2

(Y) (N)   Have you consumed anything containing caffeine in the last six hours?

(Y) (N)   Have you consumed anything high in sugar content in the last six hours?

(Y) (N)   Have you smoked cigarettes or taken any narcotics in the last six hours?

(Y) (N)   Have you engaged in any sexual activities today?

(Y) (N)   Have you watched any interesting movies/television shows today?

(Y) (N)   Have you smoked marijuana in the last 24 hours?

(Y) (N)   Have you hugged anyone recently?

(Y) (N)   Have you worked out or exercised today?

(Y) (N)   Would you say that today was a particularly good day?

(Y) (N)   Do you use an umbrella when it rains?

(Y) (N)   Do you use a toaster regularly?

How would you describe an innovative product?

Do you have any general comments on humor production or idea generation?

Appendix D
Example of a Cartoon Caption Test form

id number _____

Please start a new line for each caption.
Continue on back if needed.

_____

_____

_____

_____

_____

_____

_____

_____

_____