HumorTools: A Microtask Workflow for Writing News Satire

Lydia B. Chilton
University of Washington
hmslydia@cs.uw.edu

James A. Landay
Stanford Universtiy
landay@cs.stanford.edu

Daniel S. Weld
University of Washington
weld@cs.uw.edu

ABSTRACT

Humor is a highly-valued human skill — a sign of intelligence and creativity. Humor creation is a long-standing problem in Artificial Intelligence, because it does not easily decompose and it cannot readily be defined or detected; indeed, many humans cannot readily create jokes. However, in our survey of advice from professional comedians, we found evidence that the humor-generation process can be described. Based on this survey, we performed an analysis of news satire from *The Onion* and decomposed the process of humor creation into seven microtasks. We then developed a workflow, inspired by the design literature, that invokes these microtasks in a novel, dynamic manner. To evaluate our microtasks and workflow, we recruited 20 people, finding that the 85% of them found the workflow made their process more methodical and the microtasks enabled them to make a wider variety of jokes.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Human Factors, Design

Keywords

Microtasks, Workflow, Crowdsourcing, Humor, Design

1. INTRODUCTION

Humor is a highly valued human skill. It is a sign of intelligence and creativity, and it drives much of the entertainment industry. As an intellectual endeavor it has been a mystery for thousands of years — approached by illustrious Western thinkers such as Plato, Kant, Freud, and by modern philosophers such as Daniel Dennett [20].

Furthermore, as a type of creativity, the generation of humor presents one of the greatest challenges for artificial intelligence, requiring language abilities and world knowledge

that computers do not yet have. Indeed, many humans have trouble creating jokes on demand.

Humor-generation is challenging, in part, because there is no obvious decomposition. To approach the problem we take inspiration from philosophers' and linguists' descriptive models of humor, as well as the advice of comedians. For example, the Benign Violation Theory of humor says that humor violates our expectations [24]. This is insightful yet difficult to apply computationally. Similarly, the advice of comedians, such as introducing metaphors from seemingly unrelated domains, is too abstract to utilize directly.

In order to make progress towards the fully-automated generation of humor, we performed an analysis of 330 satirical jokes from *The Onion*, using the vocabulary and frameworks presented in the humor literature. This analysis allowed us to distill humor-creation into seven types of interacting microtasks, ranging from the identification of entities and aspects in an input headline to articulating associations and underlying beliefs. Traditional crowdsourcing uses microtasks in simple workflows, such as *iterative improvement* or *find-fix-verify* [18, 3].However, these simple control structures are insufficient for the difficult problem of humor creation. Instead, we take our inspiration from the field of design, where we search for deeper understandings of the problem, ideate in various design spaces, select an appropriate solution model based on our understanding of the space, evaluate and iterate until satisfied with the output. We find the design framework is appropriate for this creative task and can be specified at the level of microtasks. This marks a unique approach to decomposing tasks and arranging them in a dynamic workflow. Our evaluation demonstrates that our approach is successful, guiding a wide range of participants to improved humor production. In summary, this paper makes the following contributions:

- a survey of humor research and humor writing processes described by professional comedians

- seven novel microtasks, distilled from a large-scale analysis of professional news satire, that are useful for humor creation

- an example-based, tutorial that teaches our humor using the educational theory of active learning

- a novel workflow, inspired by design principles, that dynamically schedules the microtasks to cumulatively produce news satire.

- the HumorTools implementation of our workflow and microtasks and an evaluation of its performance on

20 people, finding that the 85% of them found the workflow made their process more methodical and the microtasks enabled them to make a wider variety of jokes.

2. RELATED WORK

2.1 Computational Humor

Artificial Intelligence researchers have long studied humor as a test of computer intelligence, language understanding, and creativity, but most efforts have aimed at classifying what is and is not humorous. A 2015 effort to classify submissions to *The New Yorker* Caption Contest had a 69% success rate using deep learning with human labeling of jokes and captions to add context needed for computers to understand the jokes and images [25]. Other researchers were able to classify which knock-knock jokes were funny based on presence of word play[26]. Other efforts focused on the problem of detecting sarcasm [9] with some success.

To date there have been very few effort at computational humor generation. The JAPE punning-riddle generator [?] is one of the first attempts of computational humor. It looked at both double meaning and close rhymes and found ways to fit them into question-and-answer joke format. JAPE relies heavily on puns, a form of word play that is especially tractable for computers, because word meanings are already annotated in dictionaries and understanding them requires minimal context or world knowledge. However, word play is a small subset of humor, and some of the funniest humor does rely on world knowledge and makes fun of the general human experience.. Kiddon & Brun were able to create jokes by detecting another form of word play, double-entendres [16]. Their approach has an impressive 71% accuracy, but is in essence still just a classifier, not a general humor-generation strategy. Despite this progress on special types of humor, we believe that some amount of human input is still necessary in any method for generating a broader class of humor, that goes beyond word play.

2.2 Design and Design Patterns

Design is the field that has most successfully studied open-ended problem solving. Design's approach to problem solving integrates empathetic understanding of problems, ideation, implementation and evaluation in iterative cycles [23]. Understanding the problem is often done through observation; Newman et al. observed 11 website designers to understand their problem [21]. Informed by these studies, they built tools to support the open-ended task of website design [22]. We aim to adapt Design's practice of participant observation to humor.

Design Patterns are high-level solutions to recurring engineering problems. This includes architectural patterns [2], software engineering patterns [12], and web design patterns [11]. Design patterns are an example of how hard problems can be decomposed into reusable solutions. However design patterns are still abstract, and significant effort must be put into understanding when to apply them and how to adapt them to situations. Expert design patterns can be used to automatically generate creative outcomes such as using room layout rules, generating maps, or sequencing cuts in film [1]. This body of work collects advice from professional creators, decides which of the advice or "rules" define design spaces and which rules serve as eval-

uation functions. The approach effectively turns creativity into a computational search problem. Design patterns help narrow the search, and evaluation functions give the algorithm an objective function to maximize. SHort of a fully automated approach, expert design patterns can also be used to aid people in their own creations; for example, Motif [17] uses expert patternsto help structure home videos. We use microtasks to encode comedic design patterns and guide novices in the HumorTools workflow.

2.3 Crowdsourcing and Microtasks

Crowdsourcing pioneered the web-based distribution and coordination of microtasks to complete large projects. Microtasks are attractive because they lower the barrier to entry for human workers and because they can be added into an existing software system with a simple API-call.Various workflows have been proposed to cope with the variability of human output, including iterative improvement [18], find-find-verify[3], and suggestion-and-test[7]. Some approaches use AI techniques such expectation maximization [28, 14] and decision theoretic control [8].

Crowdsourcing has attempted to automate creativity in limited but interesting ways. Yu and Nickerson [32] crowdsourced the design of chairs by mixing ideas across users to spur innovation. Yu and Kittur [31, 30] used the crowd in a two-stage, analogy-based product idea generation. The use of analogy is exciting, but additional techniques such as deep understanding of problems and empathy are other techniques from the design literature could be added.

Crowdsourcing and microtasks have been valuable to help evaluate and understand creative artifacts. Voyant [29] and CrowdCrit [19] allow users to upload visual designs and get feedback from the crowd. In the space of humor, microtasks were used to help non-native English speakers understand humorous memes in English by annotating the memes according to Semantic Script Theory of Humor [24]. Understanding and evaluation is an important part of the design process and it is clearly a problem where getting fresh perspectives and suggestions from the crowd is useful. However, evaluation is just one step in the overall process of creating novel and useful artifacts.

Although crowdsourcing often uses controlled workflows with independent microtasks, microtasks can also be used in open-ended interfaces and they don't need to be independent. Mobi [33] is a system for crowdsourcing travel plans, subject to users' preferences and constraints. Frenzy [6] is a system that conference organizers may use to group papers into relevant sessions using microtasks such as tagging, up-voting, and categorizing. These systems show that microtasks don't need to be independent nor fully coordinated to be useful. Given goals or other gentle guidance, they can be used in open-ended systems to solve complex problems.

3. SURVEY OF HUMOR LITERATURE

To inform our design of a microtask-based, humor-creation workflow, we wanted to derive as much vocabulary, theories, and process ideas from experts in humor.

3.1 Theories of Humor

Humor has been studied for thousands of years. It has been explored by Ancient Philosophers such as Plato and Aristotle, modern philosophers like Kant and Schopenhauer, and is an active field of study in linguistics and psychol-

ogy. But by and large, these thinkers have attempted to define humor and to understand *why* things are funny, not to mechanize the humor-generation process. The three major categories of humor theories found in the literature are superiority, release, and incongruity.

- **Superiority theories** claim that humor is the result of feeling superior to somebody. Plato and Hobbes describe all humor as having a basis in insult, when the listener understands the insult, they can feel superior to the people being insulted.

- **Release theories** claim that humor is release of tension built up from the suppression of discussion in society. Freud is the the champion of these theories noting that taboo subjects are often the target of jokes particularly because of the emotional release they allow.

- **Incongruity theories** claim that humor comes from the realization of an informational anomaly. The idea dates back to Aristotle but has been the focus of most modern humor theories. Benign Violation Theory [24] claims that humor comes from detecting an incongruity between expectations and reality, as long as the reality is not overly offensive. Kant, and later Schopenhauer, claimed humor was the result of perceiving a metaphor and thus seeing something in a new way that you would have previously thought unrelated.

To unite these various perspectives, philosopher Daniel Dennett [20] proposed that feelings of release and superiority can heighten the effect of a joke, but that incongruity theory is the key. He proposes the evolutionary purpose of humor is to reward debugging of incongruous information stored in our brains such as words with two meanings and double standards in society. Incongruity theories allow theorist to understand why jokes are funny by modeling the cognitive process of listeners. For example, according to Sul [24] (and others) an incongruity is not just an instance when expectations are violated, because there are many unfunny examples when expectations. To be funny, the incongruity must also *resolve* in the listeners mind. In Sul's cognitive model of humor appreciation, a joke has a setup in which the reader predicts an expected outcome. At the endf the listener predicts the outcome, then there is no surprise and thus no laughter. If there is surprise, then there is incongruity. If the listener is confused, then the incongruity is not resolved and it is not funny. However, if the listener can find a logic that makes the (unexpected) ending follow from the set up, then it is funny.

The Semantic Script Theory of Humor (SSTH), presented by Raskin in 1985 [24], is the most famous incongruity model of humor. According to SSTH, each joke can be interpreted according to two distinct, opposing scripts. One of those scripts is usually an expectation based on the set up, and the other is often an inference based on the punchline. An important aspect of this theory is that the incongruity is often in the subtext of what is said, and not the text itself.

Raskin's General Verbal Theory of Humor (GVTH) builds upon SSTH to add five more components of humor: a logical mechanism by which the incongruity is resolved, a situation which adds concreteness to the setup of the joke, a target, a narrative strategy, and language choices such as diction and word order. This suggests creating humor is challenging, since there are criteria to satisfy on many dimensions.

3.2 Survey of Humor Advice from Experts

While many people assume that humor is a completely subconscious process that cannot be taught, numerous comedians have written books claiming that humor is learnable and that there is a conscious process for creating it. We selected five popular books by comedians, describing aspects of their humor creation process. We refer to each book by the authors' last name: Carter[4], Dean[10], Kaplan[15], Holloway[13], Vorhaus[27].

3.2.1 Structure: Setup and Punchline

Jokes have set-up and punchline — the set-up establishes expectations and the punchline violates those expectations. Setup should be *relatable*, these are easy to describe in brief and will lead listeners to make many assumptions. Punchlines violate expectations by saying something that fits with the setup, but is unexpected. This aligns well with many incongruity theories of humor. Additionally, three of the books argue that the core of a punch line is truth. In summary, a good punchline is insightful — it helps you see something new about a familiar situation or object.

3.2.2 Exploration

All the comedians agree that writing a joke requires exploration of a topic. Some explorations may lead to jokes, but many will fail, and you can't know which ones will work in advance. This aligns well with the design literature on brainstorming and ideation. Holloway explicitly mentions mind maps, referring to them as "joke webs," as a way to explore a topic. Exploration can be used at many stages of the joke writing process. Carter emphasizes its use to determine a good premise for a joke, while Dean discussion the need to consider many alternative punchlines for a given set up. Part of joke writing is constraint solving and exploration at various stages of the process is a way to find solutions to the constraint.

3.2.3 Details

Jokes, like most good writing are richer when they contain details. Details make words more vivid to an audience, draw people more deeply into the work, and increases their emotional investment. Carter and Vorhaus both express a connection between details and truth. When a comedian strives for details, she has a better sense of the actual idea she is trying to portray. Carter advises that a premise for a joke should express a detailed picture of the emotion it is trying to portray.

3.2.4 Point of View

All of the comedians reinforced the importance of point-of-view in constructing jokes. Most people naturally tell stories or jokes from their own perspective. But the material becomes more immediate when it is told from the point of view of the person immediately affected. Here is an example of a joke that has had its point of view transformed from neutral to immediate:

Before:

> Q: How many Amish does it take to screw in a light bulb?
> A: Two: one to screw in the light bulb and one to wonder what it's for.

After:

Q: How many Amish does it take to screw in a light bulb?
A: What's a light bulb?

In the second version, where the listener is assumed to be the Amish person, the punchline is much more immediate.

3.2.5 Assumptions and Inference

Jokes are a story stripped down to its essentials. When writing a joke, you rely on the listener to make assumptions. Dean's example of this is the following: *"My wife just ran off with my best friend. Boy, do I miss him."* The assumption from the setup is that he misses his wife. The inference from the punchline is that he does not miss his wife, he misses his friend. Both key facts are unstated, they are assumed. Listeners will connect the dots that you omit. However, writers often need to write ideas in long form, before deciding the what to omit.

3.2.6 Emotional Quality

Emotional quality of jokes is mentioned by two of the comedians. Carter notices that her students cannot directly think about what will be funny. Instead, they can think about things that have the emotional qualities of being weird, annoying, stupid, scary or hard and then turn it into something funny. For example, wear socks with sandals is weird, in-laws are annoying, office politics are stupid, relationships are hard, and Donald Trump is scary. Any of these ideas could be the premise for a joke.

3.2.7 Constraints

Three comedians described the writing jokes as a challenge of meeting multiple constraints. They agree that exploration is necessary to satisfy the constraints when intuition fails. However, they differ in what those constraints and exploration spaces are.

For Holloway, the constraint is connecting two concepts with a word they share in common. To meet this constraint, she explores words associated with each concept separately until she finds one word they share. She uses mind maps to brainstorm associations and then look for connection. She suggests several ways to search for word associations such as taking words out of their original context, thinking of the opposite of the word, and looking for a second meaning to a word.

For Dean, the constraint is finding an alternative but valid interpretation of the information in the setup. He solves this by exploring the assumptions made by the setup and attempts to find an alternative interpretation by making different assumptions. He also suggests considering alternative points of view. His example is: *"This morning I got up, went out and ran five miles. Never push-start your car when you're alone on a hill."* The setup asserts that he ran five miles. To write a joke, he search for assumptions in this set up. One assumption is that the reason he ran is to get exercise. His punchline states alternative reason that he ran - he was chasing his car. Dean's process intentionally creates two opposing scripts because was inspired by Raskin's humor theory (SSTH).

For Carter, the constraint is finding a setup that has three things: a topic (such as "NyQuil is strong"), an attitude ("It's scary how strong it is.") and a detailed illustration of that topic and attitude ("Once I take NyQuil I'm usually asleep in 20 minutes. It knocks me out."). That premise becomes a joke by "acting out" the illustration of the attitude. For example, "It says on the back of the NyQuil box, 'May cause drowsiness.' It should say, 'Don't make any plans.'" She solves the constraint by asking "why?" questions about attitudes and details to get to a deeper understanding of her reactions and reasons. For example, if she thinks wearing socks with sandals is weird, she asks "why is it weird?" and explores the reason behind the attitude.

3.2.8 Takeaways

The expert advice on humor writing can inform our joke writing process in many ways. It confirms that humor can be at least partially if not fully externalized. Moreover, expectation Violation theory plays a role in the joke writing process. Violating an expectation can be seen as solving a constraint. However, there are multiple possible constraint and multiple spaces to explore. In addition to high-level structural techniques, there are also low-level techniques such as using details, point of view, and identifying emotional qualities in everyday occurrences.

4. HUMOR ANALYSIS

The advice from experts gives us a good vocabulary and guiding principles for developing a humor writing process. However, it remains vague and somewhat contradictory. To develop a process, we need to observe humor writing. Unfortunately, writing humor is a cognitive process, and it is hard to observe in action. Even in speaking in person with writers at The Onion, there are parts of the process that they cannot describe because happen subconsciously.

To get around our inability to observe humor writing, we looked at examples of humor and were able to back-engineer operations that could have been performing to create it. This is an ambitious task. However, we identified a domain of professional humor that make this process possible: a section of *The Onion* called *American Voices*.

The Onion is a popular and highly regarded source of humor. Much of its content is text which is easier to analyze and decompose than video or images. Most of The Onion contains fake news headlines satirizing American culture. However, they also have a section called *American Voices* which starts with real news headlines and satirizes them by writing fake "average American" responses. Table 1 contains an example of an *American Voices* style joke.

American Voices is well-suited to our analysis because the headlines can be seen as an input to the system and the jokes can be seen as the output. The job of the analysis is to find the operations that connect the headlines to the responses. Because The Onion has three jokes per headline we know there are several paths of operations between a headline and a joke.

4.1 Finding Patterns in American Voices

We performed an analysis of 330 American Voices joke, in response to 110 real headlines. In our first attempt we studied 80 American Voices joke on paper and were unable to find definitive patterns. We selected more jokes, and used a web-based tool called Frenzy [6] to make analysis faster and easier. We found four general patterns and subgroups within some of those.

4.1.1 Pattern 1: Multiple Connections

Table 1: Example of the *American Voices* style of humor: a real news headlin, a byline, and 3 jokes — fake "average American" responses.

Real Headline	Justin Bieber Baptized In NYC Bathtub
Real Byline	Pop star Justin Bieber was baptized in a friend's bathtub this weekend after weeks of Bible study and church services, with celebrity blogs reporting that the 20-year-old sought spiritual guidance in an attempt to wash away his sins following a scandal in which videos emerged of him using racial slurs.
Joke 1	"Oh my God! Can I lick the tub?"
Joke 2	"Great, now my teenage daughter's going to be begging me for $300 so she can reaffirm her devotion to God, too."
Joke 3	"Never let it be said that Bieber's PR people aren't bringing new ideas to the table."

Table 2: Example of the Multiple Connections pattern showing four concrete or oblique connections between aspects of the headline and things mentioned in the joke.

People Bending iPhones At Apple Stores	"I can't believe people would just walk into an Apple store and start breaking things like it's a Best Buy."
People	"I can't believe **people** would..."
Apple Stores	"...walk into an **Apple store**..."
bending iPhones	"...and start **breaking things**..."
Apple Store	"...like it's a **Best Buy**."

In looking at the pairs of headlines and jokes, the first thing we noticed was that headlines and jokes are connected in multiple ways: both in concrete ways and oblique ways.

The joke in Table 2 has four connections to the headline. Two of them are concrete - connecting "people" to "people" and 'Apple Stores" to "Apple store." In contrast, the other two references are more oblique: "breaking things" is a somewhat extreme characterization of "bending iPhones," and "Best Buy" is different from "Apple Store," but they are both electronics retailers.

4.1.2   Pattern 2: Association Types

After annotating the connection between headlines and jokes, some connections seemed simple to explain and some were harder. Direct connections such as "Apple Stores" to "Apple Stores" were simple to explain. However, the connection between "Apple Stores" and "Best Buy" were more oblique.

In looking at hundreds of connections, we noticed patterns in the types of associations between the headline and joke term. For example, there were many connection pairs such as "McDonald's" /"Burger King", "UConn" /'Yale", "Facebook" /"MySpace". In these pairs, the joke contained an alternative to the aspect of the headline; "McDonald's" is an alternative to "Burger King." We named this type of association "Alternative Thing." It is one of six Association Types we found. Examples of the others are shown in Table 3.

4.1.3   Pattern 3: Belief Types

Associations explain many elements in jokes but an association is only part of a joke. Often an association is just a word, and a joke needs to be a sentence. Something beyond associations is needed. Consider this example:

Report: 'SkyMall' Magazine May End Print Edition
"Alright, how many 'Summer Savannah' Backyard Garden Lion Pedestals do I have to order to turn this thing around?"

'Summer Savannah' Pedestal is associated with 'SkyMall' Magazine. It is a detailed example of the ridiculous products they sell. The detail is funny and insulting to SkyMall but it is not a joke. For a joke we need a second connection to the headline and we need a full sentence. We create a second connection by giving the speaker a belief about this headline. If the speaker is a fan of SkyMall, he does not want them to end their print edition. Thus he is posing a solution to the problem. The text "[let's] turn this thing around" is indication of his belief that there is a solution to the problem of "Ending Print Edition". Adding this belief to the 'Summer Savannah' Pedestal association gives us the second connection and completes the sentence. We found several types of beliefs hidden in the subtext of jokes. Table 4 gives examples of 4 most prominent kinds of belief in the subtext of jokes.

4.1.4   Pattern 4: Expectation Violation Types

From the humor theory and practice we know that expectation violation play a role in humor generation and appreciation. Thus, we sought to find evidence of this in the headline and joke pairs. In the *American Voices* style of humor, the headline sets up expectations, and the average American response violates them. We found evidence of two mechanisms by which responses violated expectations from the set up: sarcasm, and finding an unexpected angle to see the headline from.

Sarcasm is when the joke makes a statement the opposite to what is meant. If the statement is exaggerated enough, compared to the listener's prior beliefs then it the listener knows not interpret the statements sarcastically, not literally. Sarcasm is not always funny, it is only funny when it exposes an underlying truth or assumption about the headline. For example:

Baskin-Robbins To Honor Veterans With 'First Class Camouflage' Ice Cream

"I look forward to placing a few of these on the graves of the fallen."

The joke asserts that ice cream is a great way to honor veterans. However, this is belief is sarcastic. Certainly in comparison with the traditional way of honoring veterans (placing flowers on graves), ice cream is an inappropriate way to honor veterans. The sarcasm suggests that Baskin-Robbins' motivations are more commercial than they are solemn. The sarcasm is being used to expose this hidden facet of the headline.

Unexpected Angles There is usually an expected way people will read a headline: some aspects or will get more attention than others, and we judge headlines from our own point of view. One source of unexpected beliefs is to take an unexpected angle on on the headline such as emphasizing an overlooked detail or finding an alternative point of

Table 3: Examples of the Association Types pattern from our analysis of American Voices humor.

Association Type	Headline	Headline Aspect	Association	Joke
Alternative Thing	People Bending iPhones At Apple Stores	Apple Store	Best Buy	"I can't believe people would just walk into an Apple store and start breaking things like it's a **Best Buy**."
Alternative Person (Point of View)	Justin Bieber Baptized In NYC Bathtub	Bieber	Bieber fan	"Oh my God! **Can I** lick the tub?"
Alternative Reaction	KFC Selling Chicken Prom Corsages	chicken prom corsage	person who would want this	"My daughter's dress would look better with something from Burger King."
Specific Example	Report: 'SkyMall' Magazine May End Print Edition	SkyMall	'Summer Savannah' Backyard Garden Lion Pedestals	"Alright, how many **'Summer Savannah' Backyard Garden Lion Pedestals** do I have to order to turn this thing around?"
Insult	Toys 'R' Us Pulls 'Breaking Bad' Action Figures From Shelves	Breaking Bad	Finale was bad	"That's okay. My little guy **hated the finale**."
Personality Flaw	George Clooney Engaged	George Clooney Engaged	obsessed, delusional stalker fan	"**He usually tells me everything**, so I doubt this is true."

view. For example, in the following headline and joke about sharks, most people's reaction will be that a surging shark population is bad and dangerous. However, from the perspective of a shark, this is great news.

> *Great White Shark Populations Surging Off East Coast*
>
> *"It's an exciting time to be a shark, that's for sure."*

By thinking of an alternative person (shark) and considering an alternative belief (that surging shark population could be good, not bad), we have an unexpected angle to see this from. This expectation violation mechanism is distinct from sarcasm. From a shark's point of view, this is genuinely a good thing.

5. SYSTEM

In this section we present HumorTools, a system we built for a microtask-based joke writing workflow. We first discuss the development methodology and lessons from early prototypes, and then discuss the implementation of the final system.

5.1 Lessons from Early Prototypes

5.1.1 Assessing the Difficulty of Writing Humor

Writing humor is considered difficult, but we wanted to watch novices write humor to understand the challenges concretely. We gave six participants 10 minutes to write jokes for 5 headlines and explained they were not being judged. Their jokes would not be read by anyone else. They all expressed intimidation at the task, particularly from not knowing where to start. Surprisingly, people were able to write something down for almost every headline. However, they expressed doubts about how funny they were. Two of

the six participants stated that their jokes all had a similar sarcastic style. None of them expressed a problem with the time constraint — either they thought of a joke right away, or moved on to potentially come back later. This indicates a lack on external process for the task, and that their natural joke writing strategy is subconscious.

From this investigation of innate humor writing abilities we learned that we novices do have innate subconscious ability to write humor. However, their abilities had limits. Giving them new spaces to explore would increase the range of jokes they could make. Giving them an external process could make the tasks more conscious and less intimidating, especially when getting started.

5.1.2 Paper Prototypes

After we identified the patterns in *American Voices*, we asked two participants to apply them. Given a headline, we gave them a worksheet to write every aspect, every reaction and reason, and at least one association for each aspect. The problem with this were obvious: the branching factor is enormous and participants did not feel closer to writing a joke simply by enumerating everything, they felt overwhelmed. We learned two things about how a workflow should be designed:

First, Exploring numerous aspects at once offers many spaces to explore, but in addition to generating too many possibilities, it also disrupts the flow of thought. People think better when their thoughts can flow from one thing to a related thing. Filling out a worksheet of all possible aspects meant constantly switching to unrelated things and did not allow thoughts and associations to mature or deepen.

Second, there is a limited degree to which you can guide people's thoughts. For example, for the headline: *Americans Expected To Spend $703 Million On Pets For Valentine's Day* we asked a participant to come up with reactions to "Valentines's Day." As experimenters we assumed this would

Table 4: Examples of the Belief Types pattern from our analysis of American Voices humor.

Belief Type	Headline	Headline Aspect	Belief	Joke
Reason Good	Baskin-Robbins To Honor Veterans With 'First Class Camouflage' Ice Cream	Honor Veterans	This is good because ice cream is a great way to honor veterans	"**I look forward to** placing a few of these on the graves of the fallen."
Reason Bad	McDonald's Testing Customizable Burgers To Compete With Chipotle	Testing an improvement to fast food	This is bad because some people prefer things the way they are	"**Whatever happened to** walking into a fast food restaurant, shouting a number, and eating whatever you were given?"
Effect	Scientists Developing Heat-Resistant Chickens To Withstand Climate Change new technology	developing chicken technology	The effect on people who own chickens is upgrading their 'technology'	"How much to **upgrade** the chicken I already got?"
Solution	Report: 'SkyMall' Magazine May End Print Edition	End Print Edition	A solution to going out of business is to buy Skymall stuff.	"Alright, how many 'Summer Savannah' Backyard Garden Lion Pedestals do I have to order to **turn this thing around**?"

be easy because it was easy for us. We hate Valentines' Day. We think it is overly commercial and makes many people feel badly about themselves. However, it became clear that our reaction was not shared by one participants - she thought Valentine's Day was fine but had many reactions to "pets" such as cats who were ungrateful to their owners. She had experiences with pets and detailed associations to offer and a workflow should allow her to follow her instincts, not force her to made associations she does not have.

These observations lead us to structure microtasks in a design-based workflow. It balances creative freedom and structure. It lets people use their knowledge set and experience to guide problem solving rather than forcing people to pursue all solutions or to be structured by an rigid workflow that does not know their capabilities.

5.1.3 Tutorial Prototypes

We performed multiple rounds iteration out our tutorial involving a total of ten people. We found that although the microtasks are simple to explain and understand, people understand them better when they do exercises that involves active steps such as writing associations. This was inspired by the ICAPP learning philosophy that active learning is more effective than passive learning [5].

Based on these early prototypes we set out to build a design-based workflow for humor writing that would allow people to use their world knowledge to the fullest. We refined our tutorial to find the clearest and smallest number of examples necessary to teach a concept. Also, since people can write jokes, our objective was not to teach people to write jokes, but to teach them to write jokes with a process rather than subconsciously. We would teach them techniques to explore wider search spaces and to ease the intimidation. Ultimately we believe that teaching people search more spaces will increase the quality of their jokes. Linus Pauling said, *"The best way to have a good idea is to have a lot of ideas."*

5.2 System Description

HumorTools is a Web app built in Meteor. The system has three components: twenty microtasks, a tutorial that teaches the microtasks by example, and a workflow for applying the microtasks.

5.2.1 Microtasks and Tutorial

HumorTools introduces twenty microtasks for joke creation. The microtasks are organized into seven major types:

1. **Aspect** (1 microtask) Given a headline, identify its individual components. This may include people, things, actions or justifications for those actions. For example, in the headline "Justin Bieber Baptized in NYC Bathtub" one aspect is "Justin Bieber", another is "Baptized."

2. **Expected Reactions** (1 microtask) Given an aspect, describe your reactions to it, in or out of context. Reactions can bad positive or negative. Positive reactions include: an aspect is good or normal, deserves praise, or will succeed. Negative reaction include: an aspect is bad or weird, deserves blame, or will fail. By describing your reactions, you will probably be describing other people's expected reactions.

3. **Expected Reasons** (1 microtask) Given an aspect and a reaction, provide a reason for that reaction. This brings detail and clarity to our instinctual reactions and finds truths deeper than surface-level reactions.

4. **Associations** (6 microtasks) Given an aspect, name an associated aspect. We teach six types of associations that we found strong evidence for in section 4.1.2. For example, alternative things, points of view, and insults.

5. **Expectation Violation Mechanisms** (2 microtasks) There are two types of expectation violation mechanisms: Sarcasm and Angle. Sarcasm is when you are given an expected reaction and expected reason that seem bad or false, you name a opposite reaction and belief that is exaggerated enough to indicate you are being sarcastic. Angle is when you are given an alternative belief or association, you name a belief that is contrary to an expected belief.

6. **Beliefs** (8 microtasks) Given an expectation violation mechanism, a belief is an expression of the mechanism. For

Figure 1: Three exercises from the microtask tutorial. The highlighted text is typed by the user. Microtasks from op to bottom: Aspect, Expected Reaction and Reason, Violation.

Figure 2: The four workflow stages and examples of the microtasks belonging to each stage.

example, if the mechanism is sarcasm, a belief is the expression of the sarcastic statement. We teach eight belief types. Example of four of them are in section 4.1.3: reason something is good or bad, reason something will succeed or fail, or a good or bad effect something will have.

7. **Evaluation** (1 microtask) Given a joke, check that it has the minimum requirement to be a joke: two connections (as described in section **??**), a violation mechanism and an unexpected belief.

In the tutorial, we teach the microtasks in the order presented here. In doing so, we show that they build on one another. By starting with a headline and successively choosing microtasks to apply, you can create a joke.

The tutorials each begin with a short paragraph describing the purpose of the microtask, show one or two examples, and present exercises where users must complete the microtask. Although there are no right or wrong answers to the exercise, we make sample answers available for users to self-assess that their answer is in the right ballpark. Examples can be seen in Figure 1.

5.2.2 Dynamic Workflow

Twenty microtasks would be overwhelming to apply without a notion of a workflow to order them. However, unlike many crowdsourcing workflows, our microtasks cannot be applied in a linear workflow or even an algorithm. Our microtasks must be applied dynamically, based on the current context and opportunities available.

The workflow we teach in HumorTools follows four design principles: Understanding the problem, Ideation, applying solution patterns, and evaluation. You typically progress through the design stages linearly, however, the decision to move between these stages is dynamic: once you understand the problem sufficiently, you can move to ideation. Once ideations have matured, solutions present themselves. After solutions are executed, they must be evaluated and iterated
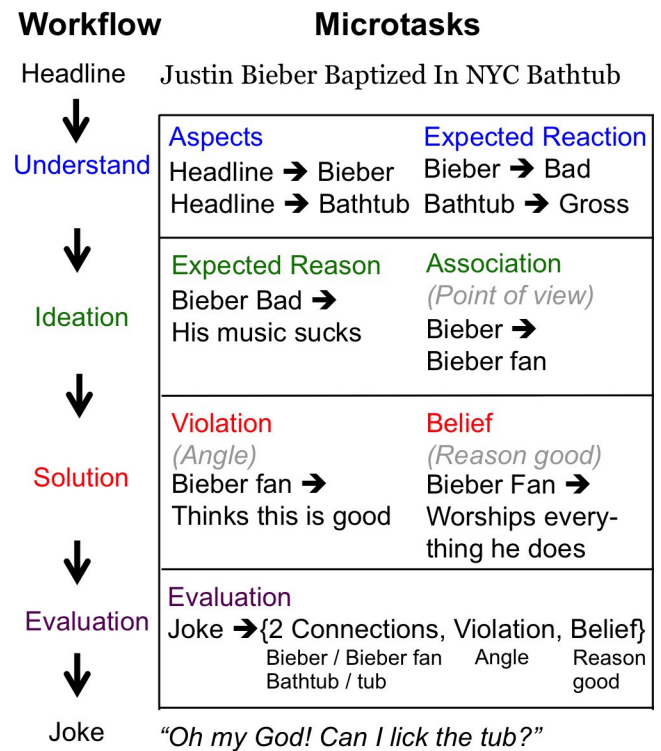
upon as necessary.

The seven types of microtasks fit into the 4 stages of the workflow as follows: Understanding comprises Aspects and Expected Reactions, Ideation comprises Expected Reasons and Associations, Solutions comprises Expectation Violation Mechanisms and Beliefs, Evaluation only contains the Evaluation microtask. Figure 2 shows this organization with a representative example of each.

Our joke writing interface is very lightweight. We give users a textarea to to write any steps they complete, a form to enter their jokes and their evaluation of the jokes, and 'cheat sheet' in the left margin to remind users of all twenty microtasks and their grouping. Because of the dynamic nature of the workflow, the freedom of a textarea was more important than the structure given by encoding the workflow in a scaffolded user interface. Once we have more data on how people apply the microtasks dynamically we could build an interface specially for this task. As it is, the tutorial clearly describes the nature of the workflow well enough for users to follow it without the guidance of an interface.

Figure 3 Shows an example of how to write jokes for a headline using this workflow. The black boxes with text represent the textareas where users write (evaluations are not shown). We have annotated the microtasks to the left of each textarea. When following the workflow, the user reads the headline, and does one of the Understanding microtasks, such as selecting an aspect such as "Justin Bieber" or "Baptized." Next, depending on context and opportunity the next microtask is selected dynamically. For the aspect "Justin Bieber" an association is made to "Bieber fan" This is an Alternative Person association and also a Point of View.

For the aspect "Baptism" an expected reason and reason are the next microtasks in this example. Both these workflows continued and ended in jokes. Two of the aspects, "NYC" and "Bathtub" never made it out of the understanding and ideation phases. This is example is simplified for length an clarity but demonstrates the dynamic nature of the workflow, and how multiple jokes can be written for the same headline depending on microtask choices.

6. EVALUATION

6.1 Study Design

To evaluate HumorTools, we wanted to see if humor novices could use the microtasks and workflow to create American Voices jokes. From our early investigations, we knew that when asked, people can write American Voices jokes, but do so by intuition. Thus, our aim was not to teach people to write jokes, or even to teach them to write better jokes (although we think this is a goal we could eventually accomplish). Our aim was to teach people to write jokes using an conscious workflow. If that is possible, we want to know which elements of the workflow are most helpful to users.

We design an web-based study lasting between 60-80 minutes that participants would complete at home. We advertised the study through mailing lists on a large college campus. Participants were paid $20 for their participation. The stages of the study were as follows:

1. Rate American Voices jokes. This gives users a warm-up and 24 examples examples of the style of humor they will be writing in. This lasts about 5 minutes.

2. Write jokes naturally. We give participants 5 minutes to write jokes for 3 headlines. This gives users a chance to experience joke writing without the tutorial so they will later be able to compare their experiences.

3. Tutorial. The tutorial on the microtasks and workflow and lasts 30-50 minutes.

4. Write Jokes with the HumorTools Workflow. We give participants 15 minutes to write jokes for the same three headlines as in stage 2. This gives users a chance to write compare their natural joke writing process to the HumorTools workflow.

5. Survey. An exit survey asking about their experience and demographics. This lasts about 5 minutes.

This study was designed to be able to elicit the qualitative differences between writing jokes with and without HumorTools. If this is successful then future work can look at the quantitative differences such as improvement in joke quality.

In our study design, the Natural joke writing step was only 5 minutes whereas the HumorTools writing steps is 15 minutes. These parameters were established in earlier prototypes. When writing joke intuitively, people do not need much time. However, in the HumorTools workflow, writing steps out takes time, so we allot more time in HumorTools to create a fair comparison.

6.2 Results

We had 20 participant (8 female, 6 male, 6 declined to say) all between the ages of 18 and 25.

6.2.1 Workflow Feedback

Of the 20 participants, 17 (85%) said that the workflow changes their joke writing process and was helpful. 3 (15%) said that the workflow was not helpful and it did not help them write jokes differently.

For the participants who said the workflow changed their process, 6 of the 17 described their the workflow make joke writing more structured, systematic or methodical. 5 of the 17 described their process before as being *"instinctual"* or they '*"didn't think about it"* or in one case their process before was to '*"[let] things sit a lot more in my head and waiting for a joke to magically pop out"* (p5). Three participants said the workflow was helpful in getting started writing jokes.

The participants who reported disliking the workflow all had different reasons. One thought it was better to just learn by example and and another said it was over-simplified and *"only for beginners."* In earlier pilots, we did find two cases of people who we selected as funny and who reported the process did nothing for them. Either they already did these or other steps in their head or they preferred to write by intuition and had a lot of success that way. We feel these are valid reasons not to use a workflow.

6.2.2 Microtask Feedback

The two microtasks participants reported most helpful were Associations (10 of 17) and Violations (10 of 17). Aspects were mentioned by 4 participants, Expected Reactions were mentioned by 4 participants, Beliefs by 2 participants (all out of 17 participants who found the microtasks useful). This wide spread is probably product of users' variety in innate humor writing abilities.

Associations were useful because they gave participants more possibilities for ideation. In their words, *"[Associations were] helpful for finding a new angle or a way to be sarcastic"* (p14) and *"[Associations] helped me think of jokes in a wider conceptual space than I previously had."* (p9) With and without the workflow this microtask was helpful: *"[Associations] generate a bunch of great ideas for a joke and makes it much easier to finish them with either the other techniques or just by thinking."* (p4)

Violations were useful because it make the actual mechanism for the joke more concrete and externalized. *"I used the violations because I knew that I had two choices – helped to narrow it down and make it feel manageable"* (p14). Particularly, of the two violation types - angle and sarcasm - angle was mentioned by 4 times participants who benefited from violations. Their natural inclination was toward sarcasm and adding angle as a second mechanism enabled them to land far more jokes. Even for people who already had natural sarcastic tendencies, other microtasks enhances their sarcastic abilities when put together in this workflow. *"Expected Reactions and Expected Beliefs [were useful] because they helped me to be sarcastic more intentionally - I could use them to make sure the jokes violated expectations"* (p14).

All the microtasks were named as useful by at least one participant. It is not surprising that Participants found different things useful. We attribute this to their difference in inherent abilities to write humor. It is hard to know which microtasks will unlock possibilities for people and which ones they either do not need or already know. In the future, being able to rapidly assess this would make teaching faster and more effective.
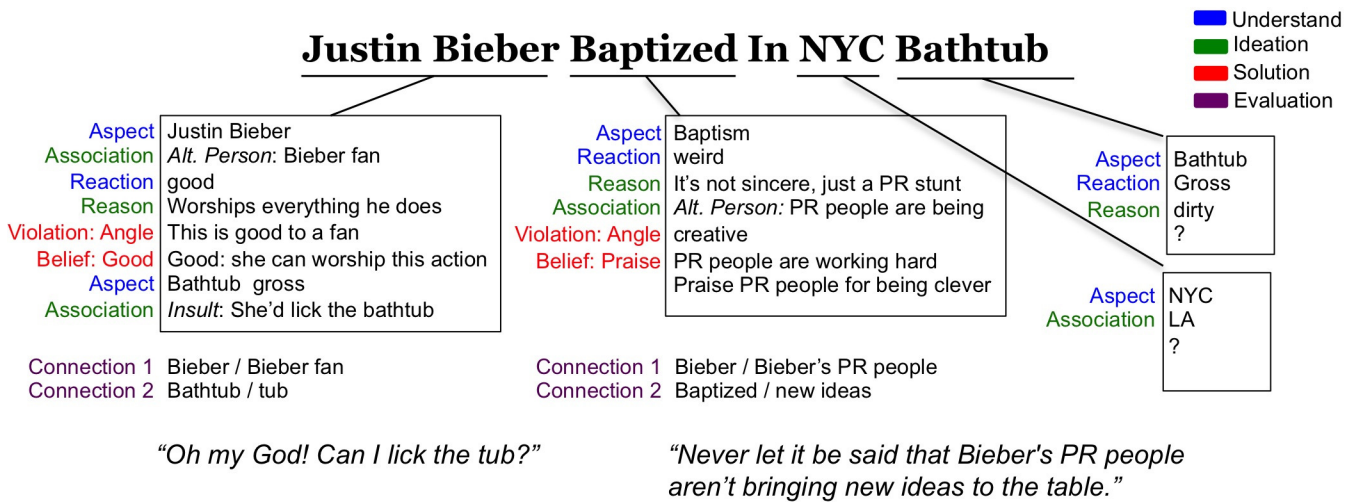
Figure 3: Example illustrating the HumorTools workflow on a headline. Two different jokes are created from the same headline by the dynamic application of microtasks.

6.2.3 Areas for Improvement

We asked participants for negative feedback. Regarding the study, four people found the tutorial too long and this probably impaired their ability to write jokes at the end. Six people found one or more of the microtasks in the tutorial obvious. One participant felt that associations did not need to be taught. Another participant found identifying aspects to be *"too analytical"* (p17). The biggest pattern is that 4 participants found Beliefs to be obvious or intuitive. We agree that Beliefs are often more more subtle. Perhaps they are obvious or perhaps more motivation or expertise would be necessary to show how powerful they can be.

It is important to note that one person felt very pressured to write a perfect joke. With any creative process, people will feel vulnerable. In future studies where we want to rate the jokes we will have to find a way to balance between the stress it causes people to have their work judged and the improvements it can bring them.

7. DISCUSSION

Our evaluation showed that we could create workflow for American Voices jokes. Could these methods generalize to other forms of humor? HumorTools is currently reliant on using headlines selected by The Onion. A first step is to determine how these headlines are being selected, then select them ourselves. Next we would be to move away from real news and start writing our own fake news. This is harder, but one possible solution is to maintain some source of inspiration for the fake news.

The bigger question is whether design-based workflows could apply to creative tasks beyond humor - writing drama, making games, designing products, writing software, or doing scientific research. Such an extension is not trivial, and largely this is a question of the power of design methodology to solve any creative task. A challenge that would have to be met is scale. We chose jokes because they are short and yet still it took us 6 months to find patterns in 330. For longer and more complex artifacts, analysis could be overwhelming. Learning from our experience with jokes, we think the process of discovering the microtasks could be streamlined

and multiple people could be involved. Collaborative data organization tools like Frenzy may need to evolve to meet these needs, and participants may need to devote days to the task. But the rewards would be worth the investment.

8. CONCLUSION AND FUTURE WORK

Humor is a highly-valued human skill and a long-standing problem in Artificial Intelligence. Humor does not easily decompose and many people cannot readily create jokes. In this paper, we introduced HumorTools - a single-person workflow that breaks the task of generating humor into microtasks that are applied dynamically according to context and opportunity. Our workflow is modeled after the design process and typically moves through four stages: understanding the problem, ideation, solution and evaluation. The application of the design process to humor was informed by a survey of humor literature including the processes of humor professional comedians, as well as our own analysis a large corpus of news satire.

In our evaluation of 20 participants, we found that the 85% of them found the workflow made their process more methodical and the microtasks enabled them to make a wider variety of jokes. This shows that the humor writing process can be externalized and we can teach people to write humor with a broader set of strategies than their subconscious humor writing processes do. In future studies we would like to show that HumorTools enable people to make more jokes than their innate process and that the HumorTools jokes are on average more funny. Ultimately, we would like to make our participants funnier than *The Onion.*

Currently, HumorTools is a single-person workflow that only allows people to build on their own ideas. Ultimately, we would like to this to be a collaborative workflow that allows people to work on other people's ideas. We believe HumorTools has the potential to be a collaborative process. Because it externalizes the workflow of humor and organizes ideas into a canonical microtasks, there are more opportunities to share information between users than there are when users write jokes subconsciously.

9. REFERENCES

[1] M. Agrawala, W. Li, and F. Berthouzoz. Design principles for visual communication. *Commun. ACM*, 54(4):60–69, Apr. 2011.

[2] C. Alexander, S. Ishikawa, M. Silverstein, M. Jacobson, I. Fiksdahl-King, and S. Angel. *A Pattern Language - Towns, Buildings, Construction.* Oxford University Press, 1977.

[3] M. S. Bernstein, G. Little, R. C. Miller, B. Hartmann, M. S. Ackerman, D. R. Karger, D. Crowell, and K. Panovich. Soylent: A word processor with a crowd inside. In *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology*, UIST '10, pages 313–322, New York, NY, USA, 2010. ACM.

[4] J. Carter. *The Comedy Bible: From Stand-up to Sitcom–The Comedy Writer's Ultimate "How To" Guide.* Touchstone, New York, New York, USA, 2001.

[5] M. T. H. Chi. Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, 1(1):73–105, 2009.

[6] L. B. Chilton, J. Kim, P. André, F. Cordeiro, J. A. Landay, D. S. Weld, S. P. Dow, R. C. Miller, and H. Zhang. Frenzy: Collaborative data organization for creating conference sessions. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, pages 1255–1264, New York, NY, USA, 2014. ACM.

[7] L. B. Chilton, G. Little, D. Edge, D. S. Weld, and J. A. Landay. Cascade: Crowdsourcing taxonomy creation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 1999–2008, New York, NY, USA, 2013. ACM.

[8] P. Dai, Mausam, and D. S. Weld. Decision-theoretic control of crowd-sourced workflows. In *AAAI10*, 2010.

[9] D. Davidov, O. Tsur, and A. Rappoport. Semi-supervised recognition of sarcastic sentences in twitter and amazon. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*, CoNLL '10, pages 107–116, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics.

[10] G. Dean. *Step by Step to Stand-Up Comedy.* Heinemann Drama, Portsmouth, New Hampshire, 2000.

[11] D. K. V. Duyne, J. Landay, and J. I. Hong. *The Design of Sites: Patterns, Principles, and Processes for Crafting a Customer-Centered Web Experience.* Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2002.

[12] E. Gamma, R. Helm, R. E. Johnson, and J. Vlissides. *Design patterns: elements of reusable object-oriented software*, volume 206. 1995.

[13] S. Holloway. *The Serious Guide to Joke Writing: How To Say Something Funny About Anything.* Bookshaker, Great Yarmouth, UK, 2010.

[14] P. G. Ipeirotis, F. Provost, and J. Wang. Quality management on amazon mechanical turk. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, HCOMP '10, pages 64–67, New York, NY, USA, 2010. ACM.

[15] S. Kaplan. *The Hidden Tools of Comedy: The Serious Business of Being Funny.* Michael Wiese Productions, Studio City, CA, 2013.

[16] C. Kiddon and Y. Brun. That's what she said: Double entendre identification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers - Volume 2*, HLT '11, pages 89–94, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics.

[17] J. Kim, M. Dontcheva, W. Li, M. S. Bernstein, and D. Steinsapir. Motif: Supporting novice creativity through expert patterns. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 1211–1220, New York, NY, USA, 2015. ACM.

[18] G. Little, L. B. Chilton, M. Goldman, and R. C. Miller. Turkit: Human computation algorithms on mechanical turk. In *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology*, UIST '10, pages 57–66, New York, NY, USA, 2010. ACM.

[19] K. Luther, A. Pavel, W. Wu, J.-l. Tolentino, M. Agrawala, B. Hartmann, and S. P. Dow. Crowdcrit: Crowdsourcing and aggregating visual design critique. In *Proceedings of the Companion Publication of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW Companion '14, pages 21–24, New York, NY, USA, 2014. ACM.

[20] D. C. D. Matthew M. Hurley. *Inside Jokes: Using Humor to Reverse-Engineer the Mind.* The MIT Press, Cambridge, MA, USA, 2011.

[21] M. W. Newman and J. A. Landay. Sitemaps, storyboards, and specifications: A sketch of web site design practice. In *Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, DIS '00, pages 263–274, New York, NY, USA, 2000. ACM.

[22] M. W. Newman, J. Lin, J. I. Hong, and J. A. Landay. Denim: An informal web site design tool inspired by observations of practice. *Hum.-Comput. Interact.*, 18(3):259–324, Sept. 2003.

[23] D. A. Norman. *The Design of Everyday Things.* Basic Books, Inc., New York, NY, USA, 2002.

[24] V. Raskin. *The Primer of Humor Research.* De Gruyter, Berlin, Germany, 2009.

[25] D. Shahaf, E. Horvitz, and R. Mankoff. Inside jokes: Identifying humorous cartoon captions. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15, pages 1065–1074, New York, NY, USA, 2015. ACM.

[26] J. M. Taylor and L. J. Mazlack. Computationally recognizing wordplay in jokes. In *In Proceedings of CogSci 2004*, 2004.

[27] J. Vorhaus. *The Comic Toolbox: How to Be Funny Even If You're Not.* Silman James Press, Los Angeles, CA, 1994.

[28] J. Whitehill, P. Ruvolo, T. Wu, J. Bergsma, and J. Movellan. Whose vote should count more: Optimal integration of labels from laberlers of unknown expertise. In *In Proc. of NIPS*, pages 2035–2043, 2009.

[29] A. Xu, S.-W. Huang, and B. Bailey. Voyant: Generating structured feedback on visual designs using a crowd of non-experts. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '14, pages 1433–1444, New York, NY, USA, 2014. ACM.

[30] L. Yu, A. Kittur, and R. E. Kraut. Distributed analogical idea generation: Inventing with crowds. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 1245–1254, New York, NY, USA, 2014. ACM.

[31] L. Yu, A. Kittur, and R. E. Kraut. Searching for analogical ideas with crowds. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, pages 1225–1234, New York, NY, USA, 2014. ACM.

[32] L. Yu and J. V. Nickerson. Cooks or cobblers?: Crowd creativity through combination. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 1393–1402, New York, NY, USA, 2011. ACM.

[33] H. Zhang, E. Law, R. Miller, K. Gajos, D. Parkes, and E. Horvitz. Human computation tasks with global constraints. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 217–226, New York, NY, USA, 2012. ACM.