Mind the Gap: Collecting Commonsense Data about Simple Experiences

Jerry S. Weltman
Louisiana State University
Baton Rouge, Louisiana
jweltm2@tigers.lsu.edu

S. S. Iyengar
Florida International University
Miami, Florida
iyengar@cis.fiu.edu

Michael Hegarty
Louisiana State University
Baton Rouge, Louisiana
mhegar1@lsu.edu

ABSTRACT
In natural language, there are many gaps between what is stated and what is understood. Speakers and listeners fill in these gaps, presumably from some life experience, but no one knows how to get this experiential data into a computer. As a first step, we have created a methodology and software interface for collecting commonsense data about simple experiences. This work is intended to form the basis of a new resource for natural language processing.

We model experience as a sequence of comic frames, annotated with the changing intentional and physical states of the characters and objects. To create an annotated experience, our software interface guides non-experts in identifying facts about experiences that humans normally take for granted. As part of this process, the system asks questions using the Socratic Method to help users notice difficult-to-articulate commonsense data. A test on ten subjects indicates that non-experts are able to produce high quality experiential data.

Author Keywords
Knowledge Acquisition; Common Sense; Large-scale Collaboration

ACM Classification Keywords
H.5.2 **[Information Interfaces and Presentation]:** User Interfaces – Natural language; I.2.6 **[Artificial Intelligence]:** Learning – Knowledge acquisition; I.2.4 **[Artificial Intelligence]:** Knowledge Representation Formalisms and Methods.

General Terms
Human Factors; Design.

INTRODUCTION
In natural language, there are many gaps between what is stated and what is understood. Consider this simple story:

> 1) *Max was on the sofa, bored, all by himself. There was a pretty vase on a little side table. He went there and picked it up. He dropped it. Crash! This was fun!*

This simple narrative contains many unstated gaps that even the youngest reader could fill in, but a computer cannot:

- Max is probably a little boy
- Max is probably in a room
- At the start of the story, Max is probably in a sitting position
- Max sees the vase before going to it
- In order to pick up the vase, Max probably walks to the side table
- To walk to the vase, Max first stands up
- Etc.

Speakers and listeners fill in these gaps, presumably from some life experience. Even very young children have collected a wealth of life experiences needed to understand simple stories. As artificial intelligence (AI) critic and philosopher Hubert Dreyfus points out, regarding early attempts to program computers to understand language:

> The programs lacked the common sense of a four-year-old, and no one knew how to give them the background knowledge necessary for understanding even the simplest stories [2].

In the many years since Dreyfus's original 1972 publication, researchers still have not been able to devise a scheme for giving computer programs the background knowledge required for understanding simple stories.

Previous attempts to collect narrative data for automatic story understanding and general natural language processing (NLP) have failed to gain traction. AI workers initially experimented with canonical abstract life scripts such as going to a restaurant or children's party [15], but these scripts are difficult to build on a large scale because they require a great deal of expertise.

Some attempts to collect life experiences via volunteer crowdsourcing [17, 18, 20] also have faltered in part because it is quite difficult for non-experts to provide narratives at a granularity suitable for commonsense

IUI'13, March 19–22, 2013, Santa Monica, CA, USA.

modeling. A more recent crowdsourcing experiment with collecting narratives [7] shows promise for inferring typical event ordering; however, the problem remains that people habitually omit what they consider to be obvious actions and states – information critical to commonsense modeling.

We propose a new methodology to help non-experts create experiential narratives. To address the natural tendency to omit details when describing an everyday experience, our methodology employs three novel techniques: 1) We model experience as a sequence of still frames, like the still images of an animation flip book. When animated, the frames expose gaps in actions; if the movement appears too abrupt, more frames need to be added. 2) We ask annotators to describe intention, emotion, location, and motion - information that is critical to a commonsense understanding of the situation. 3) We then ask annotators to explain the reason behind each description. Similar to the Socratic Method, we display the annotators' answers as a general rule, which exposes commonsense assumptions and encourages deeper explanations.

This methodology could be used as a basis for collecting detailed and coherent experiential data, which would be a fundamentally new type of resource for NLP as well as for general cognitive modeling. In the future, we envision a website that is a wiki-based collaboration, which means contributors view, discuss, and edit each other's work. The short scenes, or *experiences*, that contributors create would be open for discussion and would undergo many refinements as contributors hash out their meaning.

The overall project, called the Human eXperience Project (HXP), is a methodology and corresponding software framework that enables non-experts to create detailed narratives of simple everyday experiences. In line with the goals of McCarthy et al. outlined in [8], HXP focuses on *simple* experiences—activities and naïve mental states that one would expect to find in stories at the level of kindergarten or first grade. Concentrating on the knowledge found in children's stories helps make story understanding more tractable [8]. We know of no work that specifically focuses on collecting highly detailed child-centered experiences from non-experts. We believe that such a corpus would be a boon to statistically-oriented NLP. It would also help provide the raw data to develop new types of architectures for deep semantics and commonsense reasoning algorithms [1, 4, 10, 12, 16, 21].

We first describe the comic frames that form the basis of an experiential narrative. Then we show how users add the background, characters, and props. Next we discuss the annotation process: how the software guides the user to create statements and general rules of common sense to explain each statement. Then we describe a user test of the methodology, followed by related and future work.

COMIC FRAMES

To create an experience, a user first creates a series of comic strip frames, where each frame represents some small slice in time (e.g. one second or less). Figure 1 shows the first six frames of an experience, entitled "Max breaks the vase," based on the events in (1). The frames depict a little boy named Max sitting in the living room with nothing to do. He notices a vase and picks it up. Later, he drops it on the floor with a crash (not shown here).

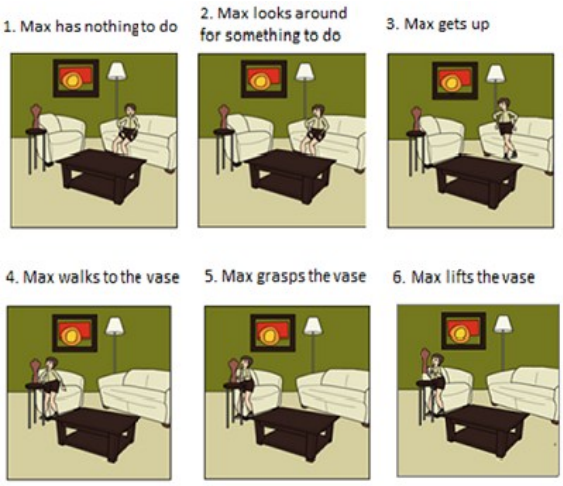

Figure 1: First six frames of "Max breaks the vase"

The format of the frames is not important at this point in the project; they can be drawings (either 2-D or 3-D), a series of photographs, or even a series of stills from a video. The only requirements are that there be multiple frames for each scene (we want a dynamic situation). For this example, we used Pixton, a free comic editing/sharing web site with reportedly hundreds of thousands of participants.[1]

When shown one after the other, the images give the illusion of animation. The goal is to make the animations appear fairly smooth. For example, an image of a boy on a sofa, followed by one of a boy next to a table, would be too abrupt; it would leave out the commonsense knowledge about how the boy gets to the table.

As an informal test, we asked two undergraduates to create some simple experiences along the style of Figure 1. After a few minutes of training, they were able to create a ten-frame narrative in about an hour.

For this paper, we will assume that the comic frames have already been created. We focus our attention on the most challenging aspect of the data collection: annotating each frame.

[1] http://pixton.com

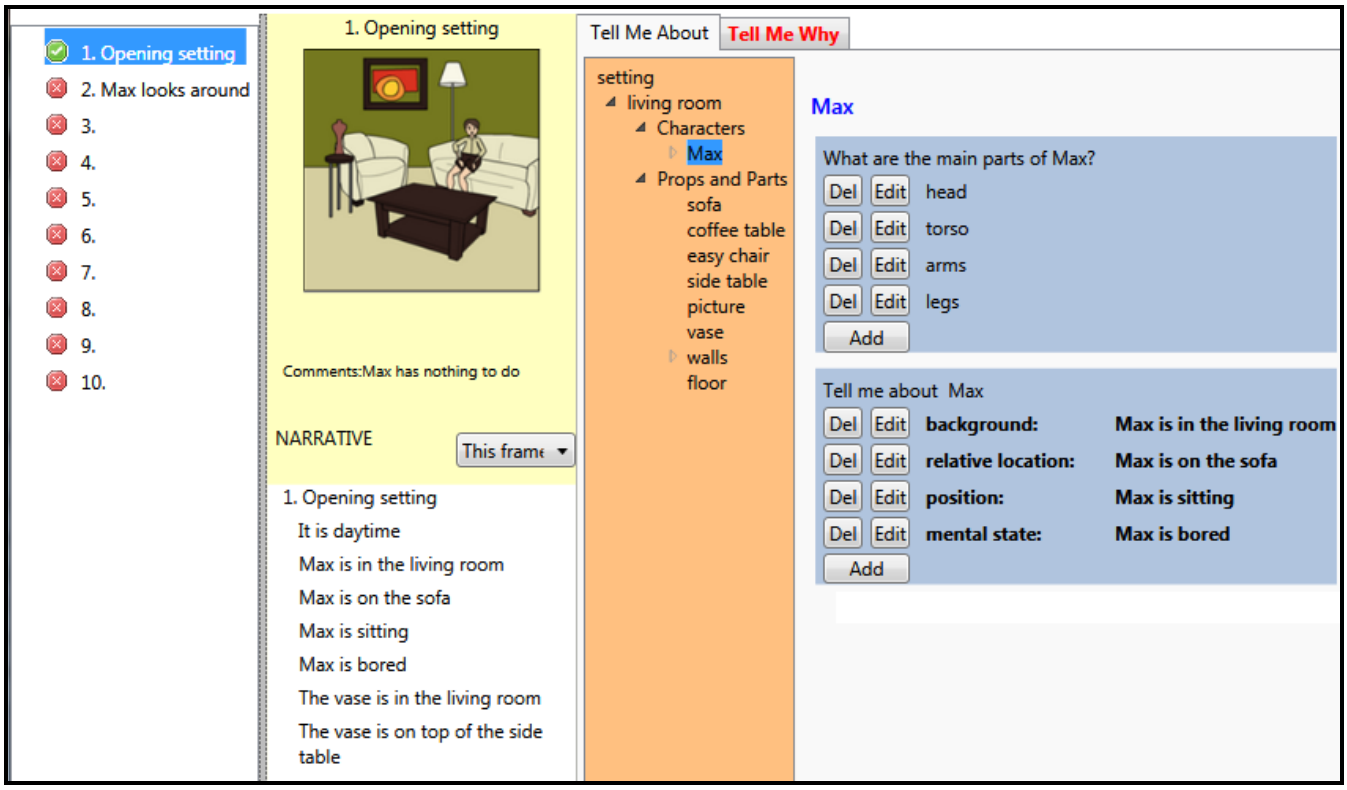

Figure 2: The opening setting after the user adds the background, character, and vase

BACKGROUND, CHARACTERS, AND PROPS

Similar to comics or movie editing software, the HXP software interface provides a set of stock background settings, characters, and props from which a contributor can populate a comic frame. Example backgrounds are a living room, kitchen, classroom, or park. Example props are a ball or vase, and example characters are a little boy and his mother.

There are currently a handful of stock backgrounds, characters, and props. The stock choices are pre-configured, but they can be edited by advanced users. (Capabilities of advanced users are currently not implemented.)

In our example experience, a user would select *living room* as the background setting. This setting comes pre-configured with many objects such as a floor, ceiling, four walls, and some furniture such as a coffee table. The user adds a vase to the frame by selecting from a list of stock props, then adds a boy from the stock characters, naming the boy *Max*. At this point, the user is prompted to input key commonsense information about Max: Max's room location, his body pose, and his mental state. Figure 2 shows the screen after the user has selected *on the sofa* for the location of Max, *sitting* for the body position, *bored* for his mental state, and *on top of the side table* for the room location of the vase.

THE ANNOTATION PROCESS

After selecting the background, character, and props, and providing key information about them, the software interface guides the user to add statements describing the characters and props. First we show an example of adding a statement using controlled natural language. Then we show how the Tell Me Why screen uses the Socratic Method to help users explain the reason for the statements.

Adding a Statement

When a user adds a statement, HXP does not parse free text, but rather structures the user input by using drop-downs and selections from controlled vocabularies. It provides feedback in natural language to show the meaning of the user's choices. This type of interface is called WYSIWYM for "What You See Is What You Mean" [14]. It controls the input so that all statements are unambiguous. All words in a statement are linked either to the WordNet [5] standard ontology or to the HXP database (for words and concepts not found in WordNet).

Figure 3 shows a screen to input "Max thinks the vase is pretty." Each of the six inputs is a drop-down choice based on the current state of the input.

In step 1, the user chooses from a list of mental states, including emotions such as *angry* and *glad* as well as complex states such as *belief* and *desire*. There are about 60 mental states currently in the system. Once the user has

chosen *think,* the system displays "Max thinks some object has some state or action" and prompts the user to choose an object from this frame.

In step 2, the user chooses among all the objects that are in the frame, including all the props, parts of props, and characters. In this example, the user chooses *vase*, and the system displays "Max thinks the vase has some state or action."

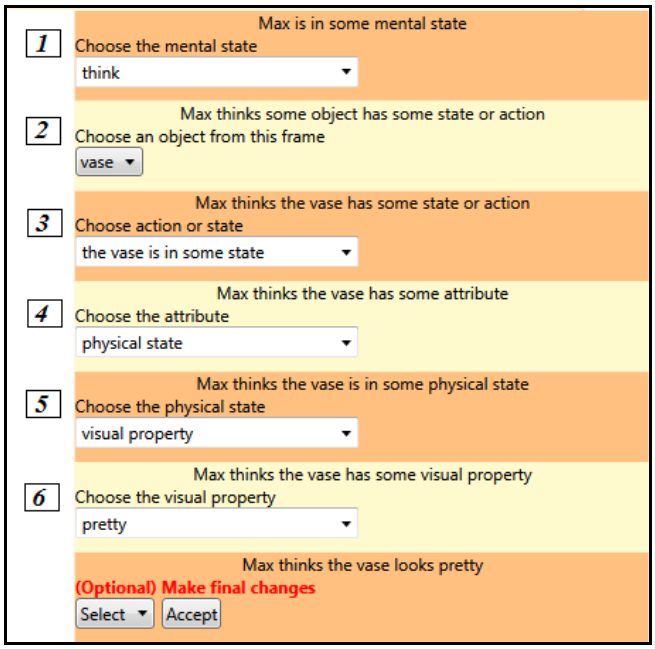

Figure 3: Input screens for the statement "Max thinks the vase looks pretty."

In step 3 the user is prompted to specify whether the vase is in some state or is doing some action. The user chooses *the vase is in some state* and continues on with the rest of the steps to drill down to *pretty*. Note, as a short cut, the user can simply type in *pretty* at any of the steps, starting at 3.

The underlying data structure representing each statement is a clause, containing a subject, predicate, optional arguments, and an optional subclause. In this example,

- Clause (subject="Max", predicate="think")
- Subclause(subject="vase", predicate="visual_attribute", argument="pretty")[2]

Different predicates require different input screens and argument structures. HXP has about a dozen general-purpose templates that control the structure of the predicate, and each predicate maps to a template. For example, there is a template for enumerated types like colors and shapes, and another template for relative location predicates like *next to*. In this example, *think* is mapped to a template that requires a subclause. Of course, there are many synonyms for *think*, such as *believe*, *imagine*, and *consider*. Users choose the most appropriate synonym, and different synonyms could map to different templates.

[2] For convenience of implementation, HXP considers predicate adjectives such as *pretty* to be binary relations.

Tell Me Why

We have seen how a user creates statements that describe the objects in a frame. Now we will see how a user creates a generalized rule that explains each statement. Going back to Figure 2, we see that the user has created seven statements, starting at the top of the Narrative section with "It is daytime." The Tell Me Why tab at the top of the figure is red, indicating that the user has not explained the reason behind these statements. Figure 4 shows the corresponding Tell Me Why screen.

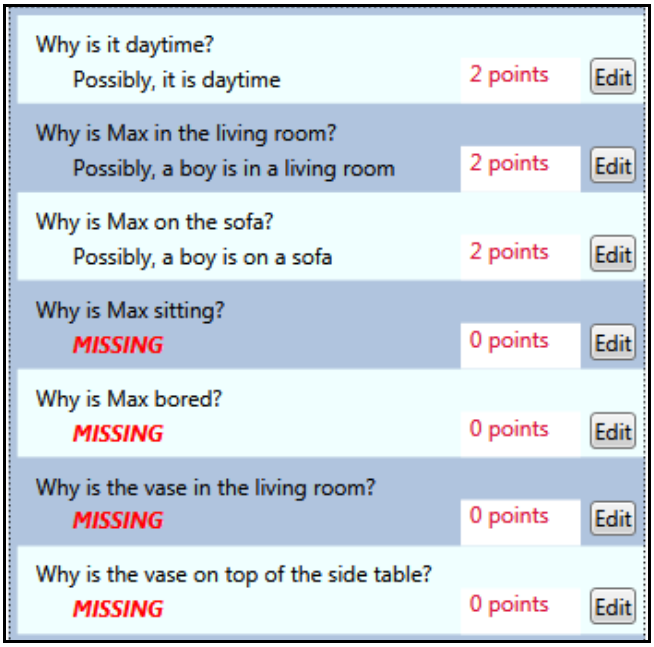

Figure 4: The Tell Me Why screen asks the user to explain each statement. The first three statements have been answered as simply "one of many possibilities." The other statements are unexplained.

As with any "Tell me why…" question, sometimes the answer is simply, "Just because I said so!" That is, the reason is too difficult to explain. In this example, there really is no good reason as to why daytime was chosen as the time of day. Therefore, the user chooses *This is just one of many possibilities* – the polite equivalent of "Just because!" Users always have the options of answering in this way, and this is perfectly fine, especially in the opening scene where the characters and setting are just being introduced.

However, even in the opening scene it is possible to provide a more informative answer to some questions. Let us look at the fourth question, "Why is Max sitting?" This

statement can be explained in terms of the previous statement "Max is on the sofa."

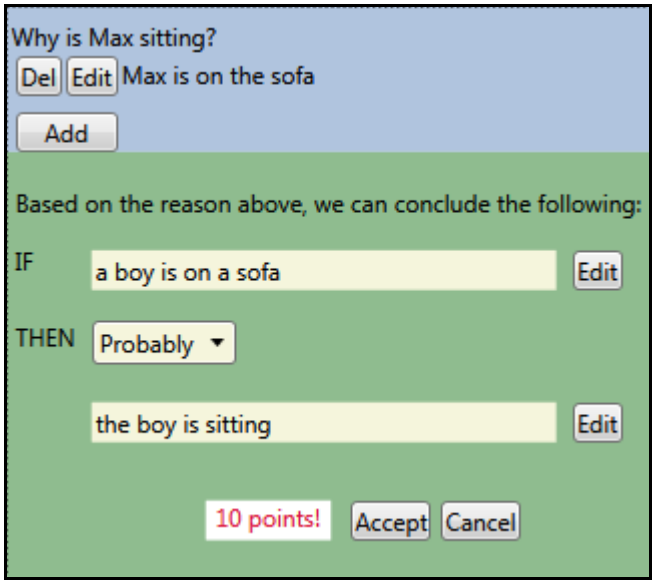

Figure 5: The user explains that we know that Max is sitting because Max is on the sofa. This explanation is then displayed as a general rule of common sense.

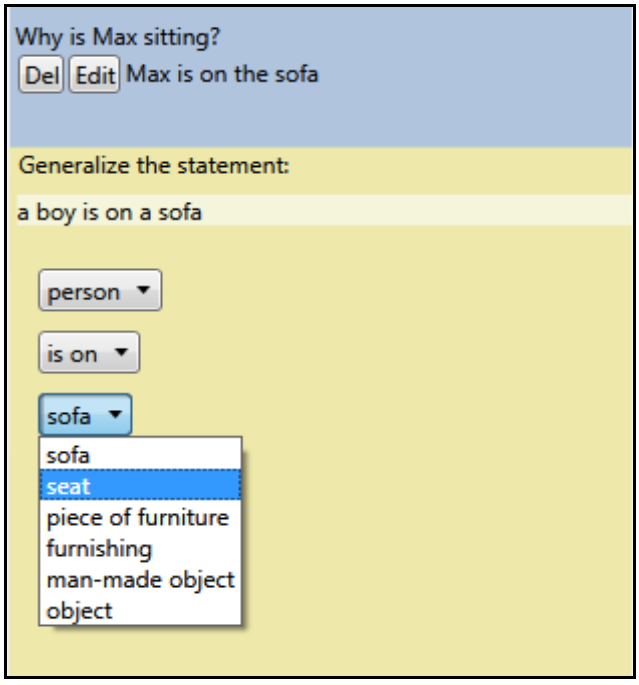

Figure 6: The user chooses hypernyms to generalize *a boy is on a sofa*. The user has already generalized *boy* to *person* and is now generalizing *sofa* to *seat*.

The relationship between being on a sofa and being in a sitting position is an unstated, but understood, rule of common sense. The HXP user interface guides the user through a series of screens to create this rule. First the user selects which of the previous statements somehow allow us to assume that Max is sitting. In this case, the user chooses the statement "Max is on the sofa." The system now restates this explanation as a general If-Then rule, as shown in Figure 5. The user has the option of editing the rule to make it more general. In this case, the user edits "a boy is on a sofa" and generalizes by choosing from the hypernyms for each noun. *Boy* has many hypernyms, including *male*, *child*, *person*, *living thing*, and *object*. Likewise, *sofa* has many hypernyms, including *seat*, *piece of furniture*, *furnishing*, *man-made object*, and *object*. The hypernyms are taken from WordNet, with a few modifications. In this example, the user's best generalization would be from *boy* to *person* and from *sofa* to *seat*.

Figure 6 shows the input screen for generalizing the statement. After generalization, the rule is:

2) IF a person is on a seat
THEN probably the person is sitting

To recap, a user first answered the question "Why is Max sitting?" by choosing a previous statement "Max is on the sofa." HXP then restated this explanation in terms of the general If-Then rule, shown in Figure 5. We have found that restating the explanation in this form exposes gaps in commonsense information and encourage users to add more detail.

Once satisfied that the rule seems to answer the question, the user proceeded to generalize it further, arriving at (2). At this point, not only has the user explained how we know that Max is sitting in this specific situation, but also the user has generated a useful rule for NLP, qualified by *probably*. In other words, given a situation where a person is on a seat, an NLP application could infer that the person is sitting, and the qualification of *probably* could be used to prioritize this rule over other possible rules.

The example illustrates the gap between what is stated and what is understood. In typical narratives, most people would assume that someone is in a state of sitting if the narrative says that the person is on a sofa.

It is instructive to look at another example to show the efficacy of this method. In Frame 2 of Figure 1, Max looks around for something to do. During the training portion of our user test, when asked why Max looks around, every one of the test subjects easily answered, "Max is bored," which is Max's mental state from Frame 1. This answer leads to rule (3), which seems correct.

3) IF a boy is bored
THEN probably the boy looks around

Moving on to Frame 3 of Figure 1, the question is much more difficult: Why does Max stand up? At first, each test

subject tried to answer again that Max is bored, generating rule (4).[3]

4) *IF a boy is bored
THEN probably the boy stands up

The subjects were generally unhappy about this rule because the mere fact that a boy is bored does not generally lead to the boy standing up. In fact, we know that Max is getting up to go to the vase. To capture this intention, one subject added the statement to Frame 2: "Max is curious about the vase" and gave this new statement as the reason for why Max gets up, generating rule (5).

5) *IF a boy is curious about a vase
THEN probably the boy stands up

When the subject saw rule (5), the subject realized that curiosity was not enough to explain why Max stands up. Other subjects added statements about Max's intention, such as "Max desires to play with the vase" or "Max desires to examine the vase." But even these statements were judged to be insufficient as soon as they were presented as, shown in rule (6).

6) *IF a boy desires to examine a vase
THEN probably the boy stands up

In order for a rule to make sense, the subjects had to add statements indicating that the vase is not near Max, and that Max is sitting, generating rule (7). [4]

7) IF a boy is sitting
AND the boy desires to examine a vase
AND the boy is not near the vase
THEN probably the boy stands up

When the subjects viewed rule (7), it seemed right; it seemed to reasonably explain why Max stands up. Once the subjects felt that the rule reasonably explains why Max stands up, they generalized it to rule (8).

8) IF a person is sitting
AND the person desires to examine an object
AND the person is not near the object
THEN probably the person stands up

Generalized rules like (8) allow specific situations to be compared to a class of situations – very useful for commonsense modeling. But the process of creating the rules itself is also useful for commonsense modeling because it leads users to add more details. Of course, these additional details spawn even more statements. Why does Max desire to examine the vase? It is because Max sees the vase and perhaps Max thinks the vase is pretty. It is important to note, however, that the user can stop the Tell Me Why cycle at any time by choosing *This is just one of many possibilities*.

The confidence levels of *possibly*, *probably*, and *definitely* are deliberately non-precise. They represent what most people would assume if they were to fill in the gap between what is stated and what is understood in the context of a typical narrative.

To motivate users to add detail and general rules, we established a simple point system on each explanation. If a user takes the easy way out with the "Just because" answer, they receive the minimum points. But if they can explain a statement in terms of a previous statement, they receive more points. And if they can generalize the statement from, say, *boy* to *person*, they get even more points. Finally, the most points are obtained by increasing the confidence level from *possibly* to *probably* or to the highest level of *definitely*.

USER TEST

We conducted a test to assess the effectiveness of our methodology. There were ten subjects: five undergraduates (music, English lit., undecided, and two from biology), four professionals with at least a bachelor's degree (accounting, psychology, business computing, linguistics), and one high school student (ninth grade). Ages ranged from 15 to 54, with half below age 25. Half were male. The evaluation consisted of a one-on-one training session for about two hours, followed by a second session where we asked subjects to annotate a few frames until they had created enough statements and explanations to score 100 points (typically five statements). In the training session, we introduced the subjects to automated NLP and our goal of collecting simple life experiences. Then we went through the first three frames of "Max breaks the vase," creating statements and explaining them with general rules. In each case, the subjects breezed through the first two frames, but as soon as they encountered "Why does Max stand up?" in the third frame, they were stuck. We spent the majority of the training talking them through the process of filling in missing information and generating relevant rules. Consequently, all of the subjects were interactively guided to produce a rule similar to Rule (8) during the training.

In the second session, we asked subjects to drive the input, specifying the statements and rules. These sessions were always with individual subjects with no collaboration. Since this was an evaluation of the methodology, not the software per se, we handled the mouse and keyboard to relieve the subjects of worrying about screen navigation. However, one subject preferred to do the navigation.

[3] The (*) in rule (4) indicates the subjects found the rule unacceptable and subsequently changed it.

[4] One test subject added a separate mental state "Max desires to walk to the vase" with a similar explanation.

1. Opening setting
   - It is daytime.
   - Max is in the living room.
   - Max is on the sofa.
   - Max is sitting.
   - Max is bored.
   - The vase is in the living room.
   - The vase is on the side table.
   - Max is naughty.
   - The side table is on the floor.

2. Max looks around the living room
   - Max sees the vase.
   - Max desires to be having fun.
   - Max desires to break the vase.
   - Max desires to drop the vase.
   - Max desires to walk to the vase.

3. Max stands up
   - Max is standing.
   - Max is next to the sofa.

4. Max walks to the side table
   - Max is near the side table.
   - Max desires to pick up the vase.

5. Max grasps the vase
   - Max is in contact with the vase.
   - The arms are in contact with the vase.

6. Max picks up the vase
   - The vase is not on the side table.
   - The vase is over the side table.
   - Max desires to turn with the vase.
   - Max desires to turn from the side table.

Figure 7: User annotations from the first six frames. The numbered statements are frame captions, representing actions. The statements below each caption result from the action.

Two of the ten subjects were able to create statements, but were confused about how to create general rules and did not complete the test. The remaining eight created an average of 11.6 statements and explained them with general rules. It took about 1-1.5 hours to produce these statements, as subjects were still getting used to the methodology. Four of these eight found the process doable, but tedious and difficult. They were happy to do the minimum and finish, each creating about five statements. However, the other four subjects said that the program was cool and "nerdy fun." They produced an average of 20 statements, with the maximum of 25. All four voluntarily continued until all frames were annotated, adding about an additional half hour to the sessions. Two of them were motivated by the points awarded to each statement, and one in particular asked what others had done and made sure to double it.

Evaluating the data

Figure 7 shows the statements from the first six frames created by one of the more prolific subjects. The statements are rich in detail, capturing intention, emotion, location, and motion. Each of the numbered statements is a frame caption. Except for the Opening setting, the captions are actions and the statements below them are states that somehow result from those actions.

All of the captions and statements were created by the user. However, we directed the user in the first three frames as part of the training session. While space does not permit us to display all of the rules that were created for these frames, Figure 8, which shows the explanations for the statements in Frame 4, represents the type of rules collected.

The first line in Figure 8 begins an explanation of the caption of Frame 4: Why does Max walk to the side table? The explanation comprises three previous statements, which are generalized. That is, *object* stands for *vase*, *person* stands for *Max*, and *table* stands for *side table*.

Why does Max walk to the side table?
- IF an object is on a table
- AND a person desires to walk to the object
- AND the person is standing
- Probably, the person walks to the table

Why is Max near the side table?
- IF a living thing goes to an object
- Definitely, the living thing is near the object

Why does Max desire to pick up the vase?
- IF a person is near a side table
- AND a jar is on the side table
- AND the person desires to drop the jar
- Probably, the person desires to pick up the jar

Figure 8: Rules created to explain the statements in Frame 4.

The second block in Figure 8 begins the explanation of a state: Why is Max is near the side table? The explanation comprises just one previous statement, the caption itself, generalized so that *living thing* now stands for *Max*, *goes* stands for *walks*, and *object* stands for *side table*.

The third block in Figure 8 begins the explanation of the state: Why does Max desire to pick up the vase? The explanation comprises generalizations of three previous statements: Max is near the side table, the vase (generalized to *jar*) is on the side table, and Max desires to pick up the vase.

Each of the three rules in Figure 8 is excellent; they adequately explain the corresponding frame statement, and they are nicely generalized so that they may apply to many situations. The first and third rules are justified in having a confidence of *probably*, because they assert what usually one would expect given the statements in the antecedent. In contrast, the second rule has a confidence of *definitely*, which is appropriate because in the context of a typical narrative, if the text says that some living thing goes to some object, then an NLP application would almost always be correct in subsequently placing the living thing near that object.

It is interesting to note how each subject's interpretations are slightly different. In two of the interpretations, Max drops the vase because it is too heavy. In another, Max is a naughty boy and desires the break the vase from the onset. In another he is angry because he is being punished. There is no expected "true interpretation." We asked only that subjects come up with reasonable statements and explanations.

We instructed the test subjects to produce rules with a confidence of at least *probably* if they could come up with a justified explanation, and all the subjects tried hard to do so. We collected 95 rules with a confidence of either *probably* or *definitely*.

To rate each rule, we used a panel of three judges, consisting of the first author and two of the more enthusiastic test subjects. Each judge independently rated all the rules as either acceptable or unacceptable. An unacceptable rule had the confidence level too high or was missing at least one critical explanation. Rule (9) below is unacceptable because the object could be something that does not normally break when dropped. Rule (10) is unacceptable because, just because a person desires to do something, we cannot say that the person will definitely do it.[5]

9) IF a person picks up an object
AND the object falls
THEN probably the object breaks

10) IF a person desires to walk to an object
AND the object is on top of a table
THEN definitely the person walks to the table

The judges found 87% of the rules to be acceptable. The opinions were unanimous on 75% of the rules, and we took the majority opinion on the remaining 25%.

Discussion

In a ten-person user evaluation, eight people were able to contribute high quality, detailed data using this methodology. Of those eight, half found the process tedious and difficult, but half found it to be challenging and fun. We are encouraged by the results that 87% of the user-built rules were acceptable, particularly because we purposely did not choose computer science majors or AI fans. Crucially, subjects could easily understand the annotations of others and found it enjoyable to identify others' problems, as attested by the fact that two judges were test subjects themselves and had no previous experience with this process.

Clearly it is unrealistic to think that we can recruit armies of casual volunteers to use this framework for collecting experiential data. But with improvements in training and a wiki format where annotators view and discuss each other's work, we believe we may be able to tap into that small percentage of the vast web population who would enjoy collaborating on this difficult task.

RELATED WORK

There are a few closely related projects for collecting narrative data. Scheherazade [3] seeks to create a bank of annotated stories to advance text understanding and narratology. StoryNet [18] and ComicKit [20], both associated with MIT's Open Mind Common Sense (OMCS) project [19], attempted to collect stories specifically for commonsense reasoning. The next two subsections describe these projects and how they relate to HXP.

Scheherazade

The Scheherazade system (henceforth SCH) is a tool for annotating stories. It is intended to be used to create a corpus of annotated texts for automated narrative analysis. HXP shares many attributes of SCH. Both systems represent intentions and goals with causal links. Both systems specifically identify actors, locations, props, and narrative time slices (called "story points" in SCH). Finally, they both use a WYSIWYM user interface to guide the user input with minimum parsing. SCH allows non-experts to paraphrase general text, and it succeeds marvelously. Indeed, SCH is not merely a proof of concept; it is a mature, ready-to-use product. Nevertheless, SCH focuses on narratological issues of plot patterns and story structure rather than on collecting commonsense data. In contrast, HXP goes after difficult-to-articulate assumptions we make when we read a text.

OMCS Projects

There have been several attempts to use narrative structures for commonsense reasoning. Schank, Minsky, and colleagues introduced the concept of scripts and frames in the 1970s [15, 11]. Scripts are abstractions of event sequences based on many concrete experiences. Mueller created a database of scripts specifically for story understanding but noted the difficulty and tedium involved with trying to create them [13]. Creating a master script to generalize an activity is extremely difficult. Indeed, it

[5] We thank one of our test subjects for suggesting this guideline: an agent's intentions or desires are never enough to warrant a confidence level of *definitely*.

requires a great deal of expertise to find the quintessential sequence of events and roles that make up a generalized *type* of activity such as going to a restaurant.

On the other hand, almost anyone can describe what happens in a particular experience at a particular time and place. As Singh and colleagues point out in their motivation for OMCS's StoryNet, "it may be easier to tell and explain a specific story, which focuses the user on a specific set of characters, objects, and events, and their relationships, than to ask them to make a general rule-based theory in the abstract of some domain" [18]. One of the reported ideas to collect stories was to have people describe their own life stories. But this proved to be too unstructured. Another idea was to make use of the OMCS data to offer an easy-to-use interface for creating structured stories. Using a simple drag and drop interface, users were to take haphazard statements from the OMCS database and put them together into a more or less coherent narrative. The following story is a combination of seven statements:

> "I travel to an airport. I board a plane. I fly in an airplane. I put on safety equipment. I open a door. I see a cloud. I jump out of an airplane."

In exchange for the easy drag and drop input, the stories were not very engaging and therefore were probably less attractive for others to read.

Another OMCS-associated experimental project was called ComicKit. It offered a comic strip interface for telling stories. Figure 9 shows an excerpt from a ComicKit. It is a story about Alice, who wakes up depressed and decides to go on a walk.

Figure 9: ComicKit Story

The comic strip idea is a good way to engage users, as indicated by ComicKit's user test, where users reportedly reported a high degree of enjoyment. But ComicKit does not address one of the basic problems inherent with comic narratives: the stories require a lot of common sense to understand what happens in the space between the comic panels [9]. Although the comic strip format reportedly was fun for users, the lack of constraints on content and captions resulted in stories that were difficult for automated analysis.

Influenced by OMCS, Li and colleagues asked minimally paid workers on Amazon's Mechanical Turk to input natural language sentences describing typical life scenarios [7]. The project's workers were instructed to use fixed actor roles, with proper names, and simple one-verb sentences. The resulting narratives were amenable for parsing, and the results indicated that automatic analysis can produce graphs of ordered events. Unlike HXP however, Li et al.'s events are not constrained to small actions and are not causally explained. Therefore, the narratives tend to leave out a great deal of detail. The example narrative from this begins with two events: a) John drives to the restaurant and b) John stands in line. This level of granularity omits the detail of John's getting in the car, traveling along a road, parking the car, getting out, or going in the restaurant. It also leaves out commonsense information about why someone would go to a restaurant, use the car, stand in line, etc. In sum, Li's narrative data is valuable, but it does not attempt to represent the contextual details like HXP.

We agree with Singh and his colleagues that collecting specific experiences should be easier than formulating generalized scripts. The challenge is providing a framework to help non-experts contribute useful data. Our comic frames are limited in time and space to small actions appropriate for animation. The frames are richly populated with realistic background settings and props, which helps contributors show how objects change from one frame to the next. Although there is a lot of content freedom, the input methodology creates highly structured data.

For the initial stages of the project, we are focusing on child experiences, which narrows the narratives to concrete actions closely tied to basic needs and naïve desires. Although the HXP methodology is not specific to child-centered experiences, we find that non-experts have an easier time when they are instructed to annotate scenes at the level of a kindergarten or first-grade reader.

To our knowledge, there have been no successful attempts to use *non-experts* to capture the commonsense detail that HXP targets: narratives, richly described with mundane, commonsense detail that other projects normally omit, accompanied by general commonsense rules. Creating these narratives is time-consuming, but by making it possible for non-experts to do the work, the intellectually laborious task of articulating life experiences can be spread among many people.

FUTURE WORK

As a result of the user test, we identified several places where the software could be improved to encourage more detailed statements. The interface should explicitly prompt

users to state the intentions/goals of the characters. Currently, it just prompts for the mental state. Also, the interface should remind users to check what states have changed from one frame to the next. Finally, we were hoping to get more statements about important sensory states, such as seeing the vase, touching the vase, and hearing the crash. We may have to add an explicit prompt for sensory states.

Our controlled natural language constructions are adequate for representing many simple experiences, but many more structures are needed. We need to add the capability of representing adverbs, comparisons, time durations, abstract concepts, group behavior, quantification, and simple dialog. Furthermore, to take the project from proof-of-concept to a fully functioning wiki collaboration, we need to implement the software as a thin client (i.e., access the program via Internet server) and add a host of features to make it more like a social networking site.

We are currently working to demonstrate the quality of the collected data by showing its inference potential. We will use forward chaining to predict what else might be true from a given statement. And we will use backward chaining to fill in the gap between what is stated and what might have been the reason behind the statement. While researchers have been demonstrating inference on manually prepared data since the early days of AI, in our case, we would be showing the efficacy of data manually prepared by non-experts aided by our software interface.

We should stress, however, that the goal here is *not* to build a rule-based NLP system; such a system would not scale up to handle a large database. Rather, the goal of HXP is strictly to collect highly structured, cohesive, generalizable data about the human experience, to be used for training scalable models or investigating new AI architectures.

Of course, a significant future challenge will be to motivate workers to participate in this project. In our small user test, some users were definitely motivated by the competition to get more points by adding more detail. They also had fun looking at other people's work and making comments and refinements. We believe other users would be motivated by a specific goal, say, creating experiences that correspond to a small corpus of children's stories similar to [8].

CONCLUSION

AI researchers have long recognized the importance of using narrative structures for natural language processing. However, attempts to narrow the problem to artificial worlds or specific domains do not scale up. Furthermore, attempts to use non-experts to provide simple stories from which commonsense can be extracted have also failed because it is so difficult for non-experts to articulate knowledge that seems so obvious.

To help non-experts create valuable narrative data, we have presented a novel methodology that focuses on simple experiences. We structure scenes into small time slices, guiding annotators to describe each frame, with particular focus on intent, emotion, location, and movement. Furthermore, we apply an automated Socratic Method to each user annotation to draw out hidden assumptions that humans make about common situations. The resulting narratives are in the form of highly structured and detailed time-ordered statements. As an added benefit, each statement is supported by a generalized rule that links *specific* situations to *classes* of situations. "Max is on a sofa" is specific situation. But rule (2) about a person on a seat is an abstraction that helps identify similarities between situations.

As [17] points out, commonsense knowledge in narrative form is contextualized; it relates situations, actions and effects. Our example narrative encodes many pieces of contextual knowledge – a person that is curious about an object might pick it up, which might mean that the person first moves to be near the object. A child that is bored in a living room might intentionally break a pretty vase. NLP applications could use this contextual knowledge to generate better translations, answer questions, and do other language understanding tasks.

HXP scenes consist of agents, actions, objects, mental states, and background setting – the same properties that cognitive scientists use to model cognition in the human brain [6]. In these models, the brain abstracts from concrete experiences as it performs essential cognitive tasks such as planning and interpreting the actions of others. If concrete experiences and abstractions of experiences are critical to cognitive processing, then we believe our methodology where users create specific experiences and explain them with generalized If-Then rules will be a fundamentally new type of resource, not just for NLP, but also for general models of human cognition.

ACKNOWLEDGEMENTS

We thank Stephen H. Elmer and Margit Link-Rodrigue for their valuable suggestions and help with this project.

REFERENCES

1. Aamodt, A. and Plaza, E. Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *AI Communications*, 7, 1 (1994), 39-59.

2. Dreyfus, H.L. *What Computers "Still" Can't Do: A Critique of Artificial Reason, Revised edition*. MIT Press, Cambridge, 1992.

3. Elson, D.K. and McKeown, K.R. Building a Bank of Semantically Encoded Narratives. In *Proc. LREC 2010*.

4. Fahlman, S.E. Using Scone's Multiple-Context Mechanism to Emulate Human-Like Reasoning. In

Proc. AAAI Fall Symposium on Advances in Cognitive Systems 2011.

5. Fellbaum, C. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA., 1998.

6. Krueger, F., Barbey, A.K., and Grafman, J. The medial prefrontal cortex mediates social event knowledge. *Trends in Cognitive Sciences*, 13, 3 (March 2009), 103-109.

7. Li, B., Appling, D.S., Lee-Urban, S., and Riedl, M.O. Learning Sociocultural Knowledge via Crowdsourced Examples. In *Proc. HCOMP 2012*.

8. McCarthy, J., Minsky, M., Sloman, A., Gong, L., Lau, T., Morgenstern, Mueller, E.T., L., Riecken, D., Singh, M., and Singh, P. An architecture of diversity for commonsense reasoning. *IBM Systems Journal*, 41, 3 (2002), 530-539.

9. McCloud, S. *Understanding Comics: The Invisible Art*. Harper Perennial, New York, 1994.

10. Minsky, M. *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. Simon & Schuster, New York, 2007.

11. Minsky, M. *A Framework for Representing Knowledge*. Technical Report. Massachusetts Institute of Technology, Cambridge, MA, 1974.

12. Mueller, E.T. Modelling Space and Time in Narratives about Restaurants. *Literary and Linguistic Computing*, 22, 1 (2006), 67-84.

13. Mueller, E.T. *Natural Language Processing with ThoughtTreasure*. Signiform, New York, 1998.

14. Power, R., Scott, D., and Evans, R. What You See Is What You Meant: direct knowledge editing with natural language feedback. In *Proc. ECAI 1998*.

15. Schank, R.C. and Abelson, R. P. *Scripts, Plans, Goals, and Understanding*. Lawrence Erlbaum, Hillsdale, NJ, 1977.

16. Schubert, L.K. Turing's Dream and the Knowledge Challenge. In *Proc. AAAI 2006*, AAAI Press (2006), 1534-8.

17. Singh, P. and Barry, B. Collecting Commonsense Experiences. In *Proc. K-CAP 2003*.

18. Singh, P., Barry, B., and Liu, H. Teaching machines about everyday life. *BT Technology Journal*, 22, 4 (October 2004), 227-240.

19. Singh, P., Lin, T., Mueller, E.T., Lim, G., Perkins, T., and Zhu, W.L. Open Mind Common Sense: Knowledge Acquisition from the General Public. In *Proc. ODBASE 2002*.

20. Williams, R., Barry, B., and Singh, P. ComicKit: Acquiring Story Scripts Using Common Sense Feedback. In *Proc. IUI 2005,* ACM Press (2005), 302-304.

21. Zarri, G.P. *Representation and Management of Narrative Information : Theoretical Principles and Implementation*. Springer-Verlag, London, 2010.