

An ASP Methodology for Understanding Narratives about Stereotypical Activities

DANIELA INCLEZAN, QINGLIN ZHANG

Miami University, Oxford OH 45056, USA
(e-mail: inclezd,zhangq7@miamioh.edu)

MARCELLO BALDUCCINI

Saint Joseph's University, Philadelphia PA 19131, USA
(e-mail: marcello.balduccini@gmail.com)

ANKUSH ISRANEY

Drexel University, Philadelphia PA 19104, USA
(e-mail: avi26@drexel.edu)

submitted 1 January 2003; revised 1 January 2003; accepted 1 January 2003

Abstract

We describe an application of Answer Set Programming to the understanding of narratives about stereotypical activities, demonstrated via question answering. Substantial work in this direction was done by Erik Mueller, who modeled stereotypical activities as *scripts*. His systems were able to understand a good number of narratives, but could not process texts describing exceptional scenarios. We propose addressing this problem by using a theory of intentions developed by Blount, Gelfond, and Balduccini. We present a methodology in which we substitute scripts by *activities* (i.e., hierarchical plans associated with goals) and employ the concept of an *intentional agent* to reason about both normal and exceptional scenarios. We exemplify the application of this methodology by answering questions about a number of restaurant stories. This paper is under consideration for acceptance in TPLP.

KEYWORDS: natural language understanding, stereotypical activities, intentions

1 Introduction

This paper describes an application of Answer Set Programming to the understanding of narratives. According to Schank and Abelson (1977), stories frequently narrate episodes related to *stereotypical activities* — *sequences of actions normally performed in a certain order by one or more actors, according to cultural conventions*. One example of a stereotypical activity is dining in a restaurant with table service, in which the following actions are expected to occur: the customer enters, he is greeted by the waiter who leads him to a table, the customer sits down, reads the menu, orders some dish, the waiter brings the dish, the customer eats and then asks for the bill, the waiter places the bill on the table, the customer pays and then leaves. A story mentioning a stereotypical activity is not required to state explicitly all of the actions that are part of it, as it is assumed that the reader is capable of filling in the blanks with his own commonsense knowledge about the activity (Schank and Abelson 1977). Consider, for instance, the following narrative:

Example 1 (Scenario 1, adapted from (Mueller 2007))

Nicole went to a vegetarian restaurant. She ordered lentil soup. The waitress set the soup in the middle of the table. Nicole enjoyed the soup. She left the restaurant.

Norms indicate, for instance, that customers are expected to pay for their meal. Readers are supposed to know such conventions, and so this information is missing from the text.

Schank and Abelson (1977) introduced the concept of a *script* to model stereotypical activities: a *fixed* sequence of actions that are *always* executed in a specific order. Following these ideas, Mueller conducted substantial work on texts about stereotypical activities: news about terrorist incidents (2004) and restaurant stories (2007). In the latter he developed a system that took as an input a restaurant story, processed it using information extraction techniques, and used a commonsense knowledge base about the restaurant domain to demonstrate an understanding of the narrative by answering questions whose answers were not necessarily stated in the text. The system performed well but the rigidity of scripts did not allow for the correct processing of scenarios describing exceptions (e.g., someone else paying for the customer’s meal). To be able to handle such scenarios, all possible exceptions of a script would have to be foreseen and encoded as new scripts by the designer of the knowledge base, which is an important hurdle.

*In this paper, we propose a new representation methodology and reasoning approach, which makes it possible to answer, in both normal and exception scenarios, questions about events that did or did not take place. We overcome limitations in Mueller’s work by abandoning the rigid script-based approach. To the best of our knowledge, ours is the first scalable approach to the understanding of **exceptional** scenarios.*

Instead, we propose to view characters in stories about stereotypical activities (e.g., the customer, waiter, and cook in a restaurant scenario), as BDI agents that *intend* to perform some actions in order to achieve certain goals, but may not always need to/ be able to perform these actions as soon as intended. It is instrumental for our purpose to use a *theory of intentions* developed by Blount et al. (2015) that introduces the concept of an *activity* — a sequence of agent actions and sub-activities that are supposed to achieve a goal. The theory of intentions is written in an action language (\mathcal{AL}) and is translatable into Answer Set Prolog (ASP) (Gelfond and Lifschitz 1991). It can be easily coupled with ASP commonsense knowledge bases about actions and their effects, and with reasoning algorithms encoded in ASP or its extensions, to build executable systems. Other implementations of BDI agents exist. Some cannot be immediately integrated in executable systems (Rao and Georgeff 1991); it remains to be seen whether others (Bordini et al. 2007) may be more readily integrated in our methodology.

Blount et al. also introduced an architecture (\mathcal{AIA}) of an *intentional agent*, an agent that obeys his intentions. According to \mathcal{AIA} , at each time step, the agent observes the world, explains observations incompatible with its expectations (diagnosis), and determines what action to execute next (planning). The reasoning module implementing \mathcal{AIA} allows an agent to reason about a wide variety of scenarios, including the serendipitous achievement of its goal by exogenous actions or the realization that an active activity has no chance to achieve its goal anymore (i.e., it is futile), illustrated by the texts below.

Example 2 (Serendipity)

Nicole went to a vegetarian restaurant. She ordered lentil soup. When the waitress brought

her the soup, she told her that it was on the house. Nicole enjoyed the soup and then left. (The reader should understand that Nicole did not pay for the soup.)

Example 3 (Detecting Futile Activity)

Nicole went to a vegetarian restaurant. She sat down and wanted to order lentil soup, but it was not on the menu. (The reader should deduce that Nicole stopped her plan of eating lentil soup here.)

Example 4 (Diagnosis)

Nicole went to a vegetarian restaurant. She ordered lentil soup. The waitress brought her a miso soup instead. (The reader is supposed to produce some explanations for what may have gone wrong: either the waitress or the cook misunderstood the order.)

In contrast with *AZA*, which encodes an agent’s reasoning process about his own goals, intentions, and ways to achieve them, we need to represent the reasoning process of a (cautious) reader that learns about the actions of an intentional agent from a narrative. For instance, while an intelligent agent creates or selects its own activity to achieve a goal, in a narrative context, the reader learns about the activity that was selected by the agent from the text. As a consequence, *one of our goals in this work is to understand what parts of the AZA architecture can be adapted to our purpose of modeling the reasoning of a narrative reader and how*. The main difficulty is that stereotypical activities normally include several agents (e.g., customer, waiter, cook), not just one. We had to extend Blount *et al.*’s theory to be able to track the intentions of several agents at a time.

Our methodology can be applied to narratives about other stereotypical activities. This is just a first exploration of the subject,¹ which can be further expanded by addressing, for instance, script variations. We propose a prototypical implementation of the end-to-end system in which understanding is tested via question answering. We convert questions to logic forms to closely match their meanings in English and encode corresponding rules in ASP to retrieve answers. Mueller’s collection of training and test excerpts is proprietary, and creating a benchmark of exceptional scenarios is a laborious task. As a result, we evaluate our work on a smaller collection of texts collected from the Internet.

In what follows, the paper provides a review of related and background work. It then continues with the description of the methodology and a preliminary implementation, and their exemplification on sample scenarios. It ends with conclusions and future work.

2 Related Work

Story Understanding. An extensive review of narrative processing systems can be found in Mueller’s paper (2007). Newer systems exist, for example the logic-based systems discussed by Michael (2013) or Diakidoy *et al.* (2015), but do not focus specifically on stereotypical activities. The task we undertake here is a more difficult one because stories about stereotypical activities tend to omit more information about the events taking place compared to other texts, as such information is expected to be filled in by the reader.

Restaurant Narratives. Erik Mueller’s work is based on the hypothesis that readers of

¹ An extended abstract on this work was previously published by Incezan *et al.* (2017); in previous work by Zhang and Incezan (2017), only the customer was modeled as a goal-driven agent.

a text understand it by constructing a mental model of the narrative. Mueller’s system (2007) showed an understanding of restaurant narratives by answering questions about time and space aspects that were not necessarily mentioned explicitly in the text. His system relied on two important pieces of background knowledge: (1) a commonsense knowledge base about actions occurring in a restaurant, their effects and preconditions, encoded in Event Calculus (Shanahan 1997) and (2) a script describing a sequence of actions performed by different characters in a *normal* unfolding of a restaurant episode. The system processed English text using information extraction techniques in order to fill out slot values in a template. In particular, it detected the last action from the restaurant script that was mentioned in the text, and constructed a logic form in which the occurrence of all actions in the script up to that last one mentioned was assumed and recorded via facts. Clearly, this approach cannot be applied to exceptional cases. For instance, the last script action identifiable in the scenario in Example 2 is that Nicole left the restaurant. As a result, the reasoning problem constructed by Mueller’s system for this excerpt would state as a fact that Nicole also executed a preceding action in the script, that of paying for her meal, which would be incorrect. Mueller’s system was evaluated on text excerpts retrieved from the web or Project Gutenberg collection. Scenarios with exceptional cases were not processed correctly because of a lack of flexibility of scripts.

Activity Recognition. The task we undertake here presents some similarities to activity recognition, in that it requires observing agents and their environment in order to complete the picture about the agents’ actions and activities. However, unlike activity recognition, understanding narratives limited to a single stereotypical activity (restaurant dining) does not require identifying agents’ goals, which are always the same for each role in our case (e.g., the customer always wants to become satiated). Gabaldon (2009) performed activity recognition using a simpler theory of intentions by Baral and Gelfond (2005) that did not consider goal-driven agents. Nieves *et al.* (2013) proposed an argumentation-based approach for activity recognition, applied to activities defined as pairs of a motive and a set of goal-directed actions; in contrast, in Blount *et al.*’s work, basic actions in an activity may, but are not required to, have an associated goal. A few decades earlier, Ng and Mooney (1992) used abduction to create a plan recognition system and tested it on a collection of short narratives that included restaurant dining. However, their system cannot reason about serendipitous achievement of an agent’s goals by someone else’s actions (Example 2), nor answer questions about an agent’s intentions.

3 Preliminary: Theory of Intentions

Blount *et al.* (Blount 2013; Blount *et al.* 2015) developed a theory about the intentions of a goal-driven agent by substantially elaborating on previous work by Baral and Gelfond (2005). In their theory, each sequence of actions (i.e., plan) of an agent was associated with a goal that it was meant to achieve, and the combination of the two was called an *activity*. Activities could have nested sub-activities, and were encoded using the predicates: $activity(\mathbf{m})$ (\mathbf{m} is an activity); $goal(\mathbf{m}, \mathbf{g})$ (the goal of activity \mathbf{m} is \mathbf{g}); $length(\mathbf{m}, \mathbf{n})$ (the length of activity \mathbf{m} is \mathbf{n}); and $comp(\mathbf{m}, \mathbf{k}, \mathbf{x})$ (the \mathbf{k}^{th} component of activity \mathbf{m} is \mathbf{x} , where \mathbf{x} is either an action or a sub-activity).

The authors introduced the concept of an *intentional agent*² — one that has goals that it intends to pursue, “only attempts to perform those actions that are intended and does so without delay.” As normally done in our field, the agent is expected to possess knowledge about the changing world around it. This can be represented as a transition diagram in which nodes denote *physical* states of the world and arcs are labeled by *physically executable* actions that may take the world from one state to the other. States describe the values of relevant properties of the world, where properties are divided into fluents (those that can be changed by actions) and statics (those that cannot). To accommodate intentions and decisions of an intentional agent, Blount *et al.* expanded the traditional transition diagram with *mental* fluents and actions. Three important mental fluents in their theory are *status*(\mathbf{m}, \mathbf{k}) (\mathbf{m} is in progress if $\mathbf{k} \geq 0$; not yet started or stopped if $\mathbf{k} = -1$), *active_goal*(\mathbf{g}) (“goal \mathbf{g} is active”), and *next_action*(\mathbf{m}, \mathbf{a}) (“the next physical action to be executed as part of activity \mathbf{m} is \mathbf{a} ”). Axioms describe how the execution of physical actions affects the status of activities and sub-activities, activates (or inactivates) goals and sub-goals, and determines the selection of the next action to be executed (see (Blount 2013) for a complete list of axioms). Mental actions include *select*(\mathbf{g}) and *abandon*(\mathbf{g}) for goals, and *start*(\mathbf{m}) and *stop*(\mathbf{m}) for activities. The new transition diagram is encoded in action language \mathcal{AL} ; in what follows, we denote by \mathcal{TI} the ASP translation of the \mathcal{AL} encoding.

Additionally, Blount *et al.* developed an agent architecture \mathcal{AZA} for an *intentional* agent, implemented in CR-Prolog (Balduccini and Gelfond 2003; Balduccini 2007) – an extension of ASP. Blount (2013) adapted the agent loop proposed by Balduccini and Gelfond (2008) and outlined the control loop that governs the behavior of an intentional agent, which we reproduce in Figure 1.

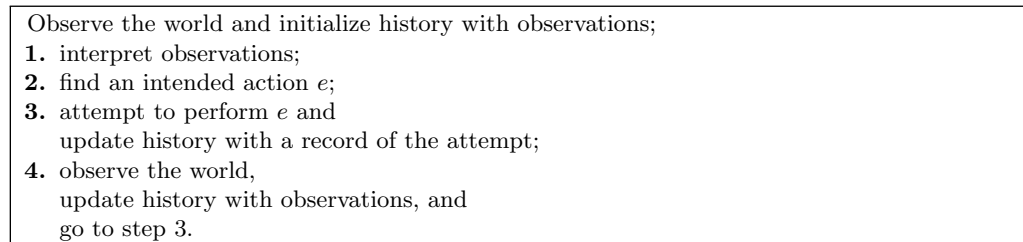


Fig. 1. \mathcal{AZA} control loop

For each step of the control loop, we provide a summary of the original description (see pages 43-44 of (Blount 2013)). In step 1, the agent uses diagnostic reasoning to explain unexpected observations, which involves determining which exogenous (i.e., non-agent) actions may have occurred without being observed. From the point of view of our

² There are similarities between Blount *et al.*’s intentional agents and BDI commitment agents (Rao and Georgeff 1991; Wooldridge 2009). To the best of our knowledge, they have not yet been studied precisely. However, a link can be drawn, at the intuitive level, as follows. If we consider the perspective that ASP formalizations are typically focused on the *beliefs* of an agent about its environment, an intentional agent can be viewed as an open-minded commitment agent. However, if the ASP formalization reflects accurately the physical reality, then it can be viewed as a single-minded commitment agent.

approach, step **2** is arguably one of the most critical. The goal of this step is to allow the agent to find an *intended action*. The following intended actions are considered:

- To continue executing an ongoing activity that is expected to achieve its goal;
 - To stop an ongoing activity whose goal is no longer active (because it has been either achieved, as in Example 2, or abandoned);
 - To stop an activity that is no longer expected to achieve its goal (as in Example 3);
- or
- To start a chosen activity that is expected to achieve its goal.

Under certain conditions, there may be no way for the agent to achieve its goal, or the agent may simply have no goal. In either case, the agent’s intended action is to wait. For the case when the agent continues executing an ongoing activity, the fluent $next_action(\mathbf{m}, \mathbf{a})$ in the theory of intentions becomes relevant as it indicates the action in activity \mathbf{m} that the agent would have to attempt next. In step **3**, the agent acts and records its attempt to perform the intended action. In the final step **4**, the agent observes the values of fluents, the result of his attempt to act from step **3**, and possibly the occurrence of some exogenous actions.

Restaurant stories require reasoning simultaneously about the intentions of multiple goal-driven agents. To accommodate for this, we added an extra argument \mathbf{ag} to associate an agent to each mental fluent and action of \mathcal{TI} (e.g., $status(\mathbf{m}, \mathbf{k})$ became $status(\mathbf{ag}, \mathbf{m}, \mathbf{k})$). We also extended \mathcal{TI} by the ASP axioms below, needed to make explicit Blount *et al.*’s assumption that an agent has only one top-level goal at a time. This restriction is important when modeling an external observer and was not fully captured previously by \mathcal{TI} . The first two axioms say that an agent cannot select a goal if it already has an active goal or if it selects another goal at the same time. The third rule says that the stopping of an activity inactivates the goals of all of its sub-activities.

$$\begin{aligned}
 impossible(select(Ag, G), I) &\leftarrow holds(active_goal(Ag, G_1), I), possible_goal(Ag, G). \\
 impossible(select(Ag, G), I) &\leftarrow occurs(select(Ag, G_1), I), possible_goal(Ag, G), \\
 &G \neq G_1. \\
 \neg holds(active_goal(Ag, G_1), I + 1) &\leftarrow goal(M_1, G_1), \\
 &holds(descendant(Ag, M_1, M), I), \\
 &occurs(stop(Ag, M), I).
 \end{aligned}$$

4 Methodology

In this section, we outline a methodology of using the theory of intentions and parts of the \mathcal{ALA} architecture to design a program that can show an understanding of stories about stereotypical activities, exemplified on restaurant stories. We distinguish between the story time line containing strictly the events mentioned in the text, and the reasoning time line corresponding to the mental model that the reader constructs. We begin with assumptions and the general methodology, on which we elaborate in the next subsections.

Assumptions. We assume that a wide coverage commonsense knowledge base (\mathcal{KB}) written in ASP is available to us and that it contains information about a large number of actions, their effects and preconditions, including actions in the stereotypical activity. How to actually build such a knowledge base is a difficult research question, but it is orthogonal to our goal. To see the first necessary steps for building such a knowledge base

see Diakidoy *et al.* (2015). In practice, in order to be able to evaluate our methodology, we have built a basic knowledge base with core information about restaurants and, whenever a scenario needed new information, we expanded the knowledge base with new actions and fluents. We operated under the assumption that all this information would be in \mathcal{KB} from the beginning. To simplify this first attempt to use a theory of intentions to reason about stereotypical activities, we assumed that there is only one customer that wants to dine, only one waiter, one cook, and one ordered dish.

Methodology. According to our methodology, for each input text t and set of questions Q , we construct a logic program $\Pi(t, Q)$ (simply $\Pi(t)$ if Q is empty). Its answer sets represent models of the narrative and answers to questions in Q . This logic program has two parts, one that is pre-defined, and another that depends on the input.

The **pre-defined part** consists of the following items:

1. The \mathcal{KB} knowledge base, with a core describing sorts, fluents, actions, and some pre-defined objects relevant to the stereotypical activity of focus;
2. The ASP theory of intentions, \mathcal{TI} ;
3. A module encoding stereotypical activities as \mathcal{TI} activities for each character; and
4. A reasoning module, encoding (i) a mapping of time points on the story time line into points on the reasoning time line; (ii) reasoning components adapted from the \mathcal{ALA} architecture to reflect a reader’s reasoning process and expected to allow reasoning about serendipitous achievement of goals, decisions to stop futile activities, and diagnosis; and (iii) a question answering component.

The **input-dependent part** (i.e., the logic form obtained by translating the English text t and questions in Q into ASP facts) consists of the following:

5. Facts defining objects mentioned in the text t as instances of relevant sorts in \mathcal{KB} ;
6. Observations about the values of fluents and the occurrences of actions at different points on the *story* time line;
7. Default information about the values of fluents in the initial situation; and
8. Facts encoding each question in Q .

4.1 The Core of the Commonsense Knowledge Base \mathcal{KB}

The core of \mathcal{KB} defines knowledge related to the restaurant environment. It includes a hierarchy of sorts with main sorts *person*, *thing*, *restaurant*, and *location*; *person* has sub-sorts *customer*, *waiter*, and *cook*; and *thing* has sub-sorts *food*, *menu*, and *bill*. In this paper, the following pre-defined instances of sorts are used: **entrance**, **kt** (kitchen), **ct** (counter), **outside**, **t** (table) are instances of *location*; **m** is a *menu*; and **b** is the customer’s bill. The core describes actions and fluents related to the restaurant environment that can be seen in Table 1, in which **t** denotes the table and we use **c** for a *customer*; **w** for a *waiter*; **ck** for a *cook*; **f** for a *food*; **r** for a *restaurant*; **t1** and **t2** for *things*; **l**, **l1** and **l2** for *locations*; **p**, **p1**, and **p2** for *persons*. We denote the agent performing each action **a** by using the static *actor(a, p)*. Each action has a unique actor, except *lead.to(w, c, t)* in which both **w** and **c** (waiter and customer) are considered actors. All fluents are *inertial* (i.e., they normally maintain their previous values unless changed by an action), except the five on the last column that are *defined-positive* fluents, i.e., their positive value is completely defined in terms of other fluents; otherwise their default value is false.

Table 1. Important actions and fluents in the restaurant-related core of \mathcal{KB}

| | | | | |
|---------|--|--|---|--|
| Actions | $go(\mathbf{c}, \mathbf{r})$ | $sit(\mathbf{c})$ | $request(\mathbf{p1}, \mathbf{t}, \mathbf{p2})$ | $stand_up(\mathbf{c})$ |
| | $greet(\mathbf{w}, \mathbf{c})$ | $pick_up(\mathbf{p}, \mathbf{t}, \mathbf{l})$ | $prepare(\mathbf{ck}, \mathbf{f})$ | $leave(\mathbf{c})$ |
| | $move(\mathbf{p}, \mathbf{l1}, \mathbf{l2})$ | $put_down(\mathbf{p}, \mathbf{t}, \mathbf{l})$ | $eat(\mathbf{c}, \mathbf{f})$ | $make_unavailable(\mathbf{f}, \mathbf{r})$ |
| | $lead_to(\mathbf{w}, \mathbf{c}, \mathbf{t})$ | $order(\mathbf{c}, \mathbf{f}, \mathbf{w})$ | $pay(\mathbf{c})$ | $interference$ |
| Fluents | $hungry(\mathbf{c})$ | $standing_by(\mathbf{p}, \mathbf{l})$ | $available(\mathbf{f}, \mathbf{r})$ | $order_transmitted(\mathbf{c})$ |
| | $open(\mathbf{r})$ | $sitting(\mathbf{c})$ | $food_prepared(\mathbf{ck}, \mathbf{f})$ | $done_with_payment(\mathbf{c})$ |
| | $at_loc(\mathbf{t}, \mathbf{l})$ | $holding(\mathbf{p}, \mathbf{t})$ | $served(\mathbf{c})$ | $satiated_and_out(\mathbf{c})$ |
| | $in(\mathbf{c}, \mathbf{r})$ | $menu_read(\mathbf{c})$ | $bill_generated(\mathbf{c})$ | $served_and_billed(\mathbf{c})$ |
| | $welcomed(\mathbf{c})$ | $informed(\mathbf{p1}, \mathbf{t}, \mathbf{p2})$ | $paid(\mathbf{b})$ | $done_with_request(\mathbf{ck}, \mathbf{w})$ |
| | | | | |

Axioms about the direct, indirect effects and preconditions of actions are encoded in ASP using standard methods (Gelfond and Kahl 2014). We show here the encoding of a direct effect of action $eat(\mathbf{c}, \mathbf{f})$ (the food is no longer on the table) and a condition that renders its execution impossible (a customer cannot eat unless the food is on the table):

$$\begin{aligned} \neg holds(at_l(F, \mathbf{t}), I + 1) &\leftarrow occurs(eat(C, F), I). \\ impossible(eat(C, F), I) &\leftarrow \neg holds(at_l(F, \mathbf{t}), I), customer(C). \end{aligned}$$

Defined fluents like $satiated_and_out$ are defined by rules like:

$$holds(satiated_and_out(C), I) \leftarrow holds(satiated(C), I), holds(at_l(C, outside), I).$$

The knowledge base \mathcal{KB} contains an exogenous action (i.e., a non-agent action) called $interference$ (see Table 1). The simultaneous occurrence of this action with $order(\mathbf{c}, \mathbf{f}, \mathbf{w})$ or $request(\mathbf{p1}, \mathbf{t}, \mathbf{p2})$ causes miscommunication, meaning that the food order or general request is not transmitted correctly. We encode this as a non-deterministic direct effect in the rules:

$$\begin{aligned} holds(informed(W, F, C), I + 1) &\leftarrow occurs(order(C, F, W), I), \\ &\quad not \neg holds(informed(W, F, C), I). \\ 1\{holds(informed(W, F_1, C), I + 1) : other_food(F_1, F)\}1 &\leftarrow \\ &\quad occurs(order(C, F, W), I), occurs(interference, I). \\ other_food(F_1, F) &\leftarrow food(F), food(F_1), F \neq F_1. \\ \neg holds(informed(W, F_1, C), I + 1) &\leftarrow holds(informed(W, F_2, C), I), F_1 \neq F_2. \end{aligned}$$

The first rule indicates that *normally* the waiter understands the customer's order correctly. An exception is when an $interference$ occurs at the same time as a customer's action of ordering some food F , which causes the waiter to understand that the customer is asking for a different food than F (second axiom above). The fourth rule says that $informed$ is a functional fluent. The \mathcal{KB} contains similar rules for the action representing all other communication acts, $request(\mathbf{p1}, \mathbf{t}, \mathbf{p2})$.

4.2 Encoding Stereotypical Activities

Stories about stereotypical activities include multiple characters with their own goals and actions. We modeled these as \mathcal{TT} activities stored in the \mathcal{KB} , as shown in Table 2 (recall that \mathbf{t} , \mathbf{m} , \mathbf{b} , \mathbf{kt} , and \mathbf{ct} stand for the table, menu, bill, kitchen, and counter respectively).

For the **customer**, we defined an activity $c_act(C, R, W, F)$ that should be read as

Table 2. Activities (and sub-activities) for the main roles in a restaurant story

| | |
|---|--|
| Customer | Name: $c_act(C, R, W, F)$ |
| | Plan: [$go(C, R)$, $lead_to(W, C, \mathbf{t})$, $sit(C)$, $c_subact_1(C, F, W)$, $eat(C, F)$, $c_subact_2(C, W)$, $stand_up(C)$, $move(C, \mathbf{t}, \mathbf{entrance})$, $leave(C)$] |
| | Goal: $satiated_and_out(C)$ |
| | Name: $c_subact_1(C, F, W)$ (sub-activity) Plan: [$pick_up(C, \mathbf{m}, \mathbf{t})$, $put_down(C, \mathbf{m}, \mathbf{t})$, $order(C, F, W)$] Goal: $order_transmitted(C)$ |
| Name: $c_subact_2(C, W)$ (sub-activity) Plan: [$request(C, \mathbf{b}, W)$, $pay(C, \mathbf{b})$] Goal: $done_with_payment(C)$ | |
| Waiter | Name: $w_act(W, C, F_1, F_2)$ |
| | Plan: [$greet(W, C)$, $lead_to(W, C, \mathbf{t})$, $move(W, \mathbf{t}, \mathbf{kt})$, $request(W, F_1, \mathbf{ck})$, $pick_up(W, F_2, \mathbf{kt})$, $move(W, \mathbf{kt}, \mathbf{t})$, $put_down(W, F_2, \mathbf{t})$, $move(W, \mathbf{t}, \mathbf{ct})$, $pick_up(W, \mathbf{b}, \mathbf{ct})$, $move(W, \mathbf{ct}, \mathbf{t})$, $put_down(W, \mathbf{b}, \mathbf{t})$] |
| | Goal: $served_and_billed(C)$ |
| Cook | Name: $ck_act(Ck, F, W)$ |
| | Plan: [$prepare(Ck, F, W)$] |
| | Goal: $done_with_request(Ck, W)$ |

“customer C goes to restaurant R where he communicates to waiter W an order for food F .” We found that modeling the customer’s activity as a nested one with sub-activities allowed reasoning about a larger number of exceptional scenarios compared to its formalization as a flat activity. We introduced sub-activities $c_subact_1(C, F, W)$ – “ C consults the menu and communicates an order for food F to W ,” and $c_subact_2(C, W)$ – “ C asks W for the bill and pays for it.” We encoded a **waiter**’s activities via objects of the form $w_act(W, C, F_1, F_2)$ – “waiter W understood that customer C ordered food F_1 and served food F_2 to him.” To allow reasoning about different types of exceptions including miscommunication, F_1 may be different from F_2 and/or the food actually ordered by C . We named the **cook**’s activities $ck_act(Ck, F, W)$ – “cook Ck prepares food F for waiter W .” Food F may not be the one ordered by the customer nor the one requested by W . We encoded activities in Table 2 via ASP rules like:

```

activity(c_act(C, R, W, F)) ← customer(C), restaurant(R), waiter(W), food(F).
comp(c_act(C, R, W, F), 1, go(C, R)) ← activity(c_act(C, R, W, F)).
...
comp(c_act(C, R, W, F), 9, leave(C)) ← activity(c_act(C, R, W, F)).
length(c_act(C, R, W, F), 9) ← activity(c_act(C, R, W, F)).
goal(c_act(C, R, W, F), satiated_and_out(C)) ← activity(c_act(C, R, W, F)).

```

and specified the actor of each activity using the predicate $actor(\mathbf{m}, \mathbf{p})$.

Default Information about Restaurant Scenarios. In relation to characters’ activities, we also define some default facts that complement the information that can be extracted from a narrative. These state that, in any restaurant scenario, the cus-

customer initially (i.e., at time step 0 on the reasoning time line) selects the goal of being *satiated_and_out*. Similarly, the waiter selects its goal *served_and_billed(C)* as soon as customer *C* arrives, and the cook selects its goal *done_with_request(Ck, W)* immediately after receiving a food request from waiter *W*. This is encoded as follows:

$$\begin{aligned} & \text{occurs}(\text{select}(C, \text{satiated_and_out}(C)), 0) \leftarrow \text{customer}(C). \\ & \text{occurs}(\text{select}(W, \text{served_and_billed}(C)), I + 1) \leftarrow \text{waiter}(W), \\ & \quad \text{holds}(\text{arrived}(C, R), I + 1), \neg \text{holds}(\text{arrived}(C, R), I). \\ & \text{occurs}(\text{select}(Ck, \text{done_with_request}(Ck, W)), I + 1) \leftarrow \text{cook}(Ck), \\ & \quad \text{holds}(\text{informed}(Ck, F, W), I + 1), \neg \text{holds}(\text{informed}(Ck, F, W), I). \end{aligned}$$

Once a goal is selected, the program assumes that each character starts one of their candidate activities that can achieve this goal. The assumption is formalized by:

$$1\{\text{occurs}(\text{start}(P, M), I + 1) : \text{actor}(M, P), \text{goal}(M, G)\}1 \leftarrow \text{occurs}(\text{select}(P, G), I).$$

4.3 Reasoning Module

We now concentrate on the reasoning module of the *ALA* architecture associated with *TI*, and determine what parts of it can be imported/ adapted and what new rules need to be added in order to capture the reasoning process of a reader of a narrative. In what follows, we denote the reader's reasoning module by *RM* and start with a few key points in its *ALA*-inspired construction:

- A reader needs to map observations about fluents and actions (i.e., a *history*) into the predicates *holds* and *occurs* used for reasoning. It also needs to perform diagnosis (just like an intentional agent would) when reading pieces of information that are unexpected, such as a waiter bringing the wrong dish. Thus *the temporal projection and diagnostic modules of ALA are imported into RM*.

- The reader needs to fill the story time line with new time points (and thus construct what we call a *reasoning time line*) to accommodate mental and physical actions not mentioned in the text. *This is achieved by adding new rules to RM*, in which we denote story vs. reasoning time steps by predicates *story_step* and *step*, respectively, and introduce predicate *map(s, i)* to say that story step *s* is mapped into reasoning step *i*:

$$\begin{aligned} & 1\{\text{map}(S, I) : \text{step}(I)\}1 \leftarrow \text{story_step}(S). \\ & \neg \text{map}(S, I) \leftarrow \text{map}(S_1, I_1), S < S_1, I \geq I_1, \text{story_step}(S), \text{step}(I). \end{aligned}$$

Information about story steps that need to be mapped into consecutive reasoning time steps, extracted from the input in the form of *next_st(s, s1)* facts, is encoded by the rule:

$$\text{map}(S_1, I + 1) \leftarrow \text{next_st}(S, S_1), \text{map}(S, I).$$

Observations about the occurrence of actions and values of fluents recorded from the text using predicates *st_hpd* and *st_obs* as described in Subsection 4.4 are translated into observations on the reasoning time line via the rules:

$$\begin{aligned} & \text{hpd}(A, V, I) \leftarrow \text{st_hpd}(A, V, S), \text{map}(S, I). \\ & \text{obs}(F, V, I) \leftarrow \text{st_obs}(F, V, S), \text{map}(S, I). \end{aligned}$$

Finally, we do not want to create time steps on the reasoning time line when no action is believed to have occurred (i.e., gaps). We encode this using the rules

$$\begin{aligned} & \text{smtg_occurs}(I) \leftarrow \text{occurs}(A, I). \\ & \leftarrow \text{last_assigned}(I), \text{step}(J), J < I, \text{not smtg_occurs}(J). \end{aligned}$$

where *last_assigned(i)* is true if *i* is the last time step on the reasoning time line that has a correspondent on the story time line:

$$\neg \text{last_assigned}(I) \leftarrow \text{map}(S, J), \text{step}(I), \text{step}(J), I < J.$$

$$\text{last_assigned}(I) \leftarrow \text{map}(S, I), \text{not } \neg \text{last_assigned}(I).$$

- A reader may be asked questions about the story. We support yes/no, when, who, and where questions related to events. A question is represented by an atom $\text{question}(q)$, where q is a term encoding the question, e.g., $\text{occur}(A)$ (“did action A occur?”), $\text{when}(A)$ (“when did A occur?”). Answers are encoded by atoms $\text{answer}(q, a)$, where a is the answer. For example, $\text{answer}(\text{occur}(\text{pay}(\text{nicole}, b)), \text{yes})$ states that the answer to question “Did Nicole pay the bill?” is yes. Rules are used for identifying and returning the answers. Due to space limitations, we only briefly illustrate the rules for $\text{question}(\text{occur}(A))$. Suppose we want to reason about the occurrence of a specific event in a story. In that case, a positive answer is returned if the reader has definite reasons for believing that the event happened. This is encoded by the rule:

$$\text{answer}(\text{occur}(A), \text{yes}) \leftarrow \text{question}(\text{occur}(A)), \text{physical_action}(A), \text{step}(I), \\ \text{occurs}(A, I).$$

Answering a definite “no” is less straightforward, as it requires ensuring that the action did not happen at *any* step:

$$\text{maybe}(A) \leftarrow \text{physical_action}(A), \text{step}(I), \text{not } \neg \text{occurs}(A, I).$$

$$\text{answer}(\text{occur}(A), \text{no}) \leftarrow \text{question}(\text{occur}(A)), \text{physical_action}(A), \text{step}(I), \\ \text{not } \text{answer}(\text{occur}(A), \text{yes}), \text{not } \text{maybe}(A).$$

The first rule states that A *may have occurred* if, for some step I , there is lack of evidence that I did not occur. The second rule yields answer “no” if there is no evidence that the action definitely occurred and no reason to believe that the action may have occurred.

Blount *et al.* considered four categories of histories and described the encoding of the corresponding agent strategies in \mathcal{ALA} . We now analyze each category separately and discuss its suitability for RM , as well as pertinent changes:

1. No goal nor activity to commit to. *In \mathcal{ALA} , the agent waits.* In stereotypical activities, agents have active pre-defined goals/ activities, so this is not relevant to RM .

2. A top-level activity is active but its goal is not. *In \mathcal{ALA} , the agent stops the activity.* Serendipitous achievement of a character’s goal by someone else’s actions can happen in narratives about stereotypical activities (see Example 2). Thus \mathcal{ALA} histories of category 2 are relevant and the corresponding rules from \mathcal{ALA} are included in RM .

3. A top-level activity and its goal are active. *In \mathcal{ALA} , the agent performs the next action, unless there are no chances for the activity to still achieve its goal (i.e., the activity is deemed futile), in which case it is stopped.* RM imports the corresponding rules for the agent strategy, but changes the definition of a futile activity, since RM captures the reasoning of an external observer, not that of the agent involved in the activity. We add default background knowledge about situations that render an activity futile to the module describing activities and sequences (see Subsection 4.2). For example, we encode the information that an activity of type $c_act(\mathbf{c}, \mathbf{r}, \mathbf{w}, \mathbf{f})$ is futile if \mathbf{r} is observed to be closed when \mathbf{c} wants to enter; if \mathbf{f} is observed to be unavailable at \mathbf{r} ; etc., via rules like:

$$\text{futile}(c_act(C, R, W, F), I) \leftarrow \text{obs}(\text{open}(R), \text{false}, I), \text{activity}(c_act(C, R, W, F)).$$

$$\text{futile}(c_act(C, R, W, F), I) \leftarrow \text{obs}(\text{available}(F, R), \text{false}, I), \\ \text{activity}(c_act(C, R, W, F)).$$

More complex rules can also be added, such as that this activity is futile if no table is available and the customer is impatient.

4. A goal is active but there is no active activity to achieve it. *The agent*

needs to find a plan (i.e., start a new activity). Our cautious reader is not expected to guess or assume the new plan that a character computes or selects, but rather determine that re-planning is required. We introduce a new mental action $replan(\mathbf{g})$ where \mathbf{g} is a goal, and add the following rule to RM :

$$occurs(replan(Ag, G), I) \leftarrow categ_4_hist(G, I), \text{ not } impossible(replan(Ag, G), I), \\ \text{ not } futile(Ag, G, I).$$

The complete set of RM rules for reasoning about activities and categories of histories is shown in Figure 2.

| | | | |
|----|--|--------------|---|
| 1 | $categ_2_hist(Ag, M, I)$ | \leftarrow | $\neg holds(minor(Ag, M), I), holds(active(Ag, M), I),$ |
| 2 | | | $goal(M, G), \neg holds(active_goal(Ag, G), I).$ |
| 3 | $categ_3_hist(Ag, M, I)$ | \leftarrow | $\neg holds(minor(Ag, M), I), holds(in_progress(Ag, M), I).$ |
| 4 | $categ_4_hist(Ag, G, I)$ | \leftarrow | $\neg holds(minor(Ag, G), I), holds(active_goal(Ag, G), I),$ |
| 5 | | | $\neg holds(in_progress(Ag, G), I).$ |
| 6 | $occurs(stop(Ag, M), I)$ | \leftarrow | $categ_2_hist(Ag, M, I).$ |
| 7 | $occurs(AA, I)$ | \leftarrow | $categ_3_hist(Ag, M, I), \text{ not } futile(Ag, M, I),$ |
| 8 | | | $\neg holds(minor(Ag, M), I), holds(in_progress(Ag, M), I),$ |
| 9 | | | $holds(next_action(Ag, M, AA), I), \text{ not } impossible(AA, I).$ |
| 10 | $occurs(stop(Ag, M), I)$ | \leftarrow | $categ_3_hist(Ag, M, I), futile(Ag, M, I), activity(M).$ |
| 11 | $mental_inertial_fluent(replanned(Ag, G))$ | \leftarrow | $agent(Ag), possible_goal(Ag, G).$ |
| 12 | $holds(replanned(Ag, G), I + 1)$ | \leftarrow | $occurs(replan(Ag, G), I).$ |
| 13 | $impossible(replan(Ag, G), I)$ | \leftarrow | $holds(replanned(Ag, G), I).$ |
| 14 | $impossible(replan(Ag, G), I)$ | \leftarrow | $occurs(start(Ag, M), I), possible_goal(Ag, G).$ |
| 15 | $occurs(replan(Ag, G), I)$ | \leftarrow | $categ_4_hist(Ag, G, I), \text{ not } impossible(replan(Ag, G), I),$ |
| 16 | | | $\text{ not } futile(Ag, G, I).$ |
| 17 | $occurs(wait(Ag), I)$ | \leftarrow | $categ_4_hist(Ag, G, I), futile(Ag, G, I).$ |
| 18 | $justified(Ag, A, I)$ | \leftarrow | $holds(next_action(Ag, M, A), I), physical_agent_action(Ag, A).$ |
| 19 | | \leftarrow | $occurs(A, I), \text{ not } justified(Ag, A, I), physical_agent_action(Ag, A).$ |

Fig. 2. RM rules for reasoning about activities and histories.

This concludes the description of the main parts of RM and that of the pre-defined parts in the logic program constructed according to our methodology.

4.4 Logic Form

Now, we describe the input-dependent part of the logic program that is constructed for every narrative. The text is translated into a logic form that contains two parts: (a) definitions of instances of relevant sorts, e.g. *customer*, *waiter*, *food*, *restaurant*; and (b) observations about the values of fluents and occurrence of actions in relation to the *story* time line. While this may be accomplished in various ways, currently we adopt an approach that comes from a combination of techniques introduced by Lierler

| |
|--|
| $r1, r2, e1$ |
| $entity(r1), entity(r2), property(r1, nicole), property(r2, \text{"vegetarian restaurant"})$ $event(e1), eventType(e1, go_01), eventArgs(e1, a1, r1), eventArgs(e1, a4, r2)$ |

Fig. 3. DRS for the sample sentence

et al. (2017) and Balduccini *et al.* (2007). In the former, the text is translated into a Discourse Representation Structure (DRS) (Kamp and Reyle 1993) by integrating the outputs of the LTH semantic role labeler and of CORENLP. While the DRS is a logical representation, it is still largely focused on linguistic elements rather than on concepts relevant to an action theory. Balduccini *et al.* (2007) bridged this gap by means of special ASP rules that carry out a final mapping step into an Object Semantic Representation (OSR) featuring the elements listed earlier in this paragraph.

As an example, let us illustrate how the sentence “Nicole went to a vegetarian restaurant” is translated to our target logic form. Following Lierler *et al.*’s approach (2017), LTH is used to identify entities and verbs and to label entities by the roles they play in the verbs as identified by the PROPBANK³ frame schemas. This results in

$$[V (go.01) went] [A1 Nicole] [A4 \text{"to a vegetarian restaurant"}]$$

where *go.01* denotes sense 1 of verb *go* from the Ontonotes Sense Groupings, *A1* is the verb’s *entity in motion*, and *A4* is its *end point*. Next, CORENLP carries out mention detection and coreference resolution, grouping the phrases from the output of LTH that denote the same object. Then, a postprocessing step occurs, which assigns unique labels to the the entities of the DRS, yielding the first two rows of the DRS from Figure 3. This step also leverages the PROPBANK frame schemas to generate the description of the events of the DRS found in the third row of Figure 3. Finally, the postprocessing step assigns time steps to the verbs according to their syntactic ordering in the passage, resulting in the final row of the Figure. *OSR rules* from (Balduccini et al. 2007) are then used to generate the final representation of the story. While at this stage the OSR rules are created manually, it is conceivable that they can be automatically generated from the PROPBANK frame schemas. A sample rule that maps verb *go.01* to the occurrence of an action is:

$$\begin{aligned}
st_hpd(go(Actor, Dest), true, S) \leftarrow \\
event(Ev), eventType(Ev, go_01), \\
eventArgs(Ev, a1, EActor), property(EActor, Actor), \\
eventArgs(Ev, a4, EDest), property(EDest, Dest).
\end{aligned}$$

The ORS rules can also introduce new constants (e.g., *cook1*) when the name of one of the entities is not given in the text. Note that, in order to distinguish between the story time line and the reasoning time line, we substitute the predicates *obs(f, v, i)* and *hpd(a, v, i)* normally used in the description of histories (Blount et al. 2015) by *st_obs(f, v, s)* (fluent *f* from the \mathcal{KB} has value *v* at time step *s* in the story time line, where *v* may be **true** or **false**) and *st_hpd(a, v, s)* (action *a* from the \mathcal{KB} was observed to have occurred if *v*

³ The PROPBANK project (Palmer et al. 2005) provides a corpus of text annotated with information about basic semantic propositions.

is **true**, or not if **v** is **false**, at time step **s** in the story time line). In addition to the observations obtained directly from the text (mentioned there explicitly), the logic form also contains *default*, commonsensical observations such as the fact that the restaurant is assumed to be open, the customer is hungry, the waiter is at the entrance, and so on.

Example 5 (Logic Form for a Text)

The text in Example 1 is thus translated into a logic form that includes the following facts in addition to the default observations:

| | |
|----------------------------------|---|
| <code>customer(nicole).</code> | <code>st_hpd(go(nicole, veg_r), true, 0).</code> |
| <code>restaurant(veg_r).</code> | <code>st_hpd(order(nicole, lentil_soup, waitress), true, 1).</code> |
| <code>food(lentil_soup).</code> | <code>st_hpd(put_down(waitress, lentil_soup, t), true, 2).</code> |
| <code>waitress(waitress).</code> | <code>st_hpd(eat(nicole, lentil_soup), true, 3).</code> |
| <code>cook(cook1).</code> | <code>st_hpd(leave(nicole), true, 4).</code> |

Questions are translated, in a similar way, to a logic form consisting of atoms of the form *question(q)*, discussed earlier. For instance, questions q1 and q2 below are linked to actions *leave(nicole)* and *pay(nicole, b)* from the *KB*, resp., yielding the logic forms:

Did Nicole leave the restaurant? Logic form: *question(occur(leave(nicole)))* (q1)

Did Nicole pay for the soup? Logic form: *question(occur(pay(nicole, b)))* (q2)

5 Methodology Application

We applied our methodology to a collection of stories describing normal scenarios and different types of exceptional scenarios. In this section, we exemplify our outcomes on a few illustrative stories. We test understanding via questions about the occurrence of an event, which may not be explicitly mentioned in the text. Such questions are problematic for systems based on statistical methods. Given that we use a logic approach, based on ASP and its extensions, the correctness of the answers provided by our methodology can be proven for certain classes of text-question pairs via methods similar to those employed by Todorova and Gelfond (2012). In the future, we plan to perform a more formal evaluation on a larger corpus, possibly based on recent story benchmarks (e.g., (Richardson et al. 2013; Mostafazadeh et al. 2016)). Mueller’s corpora are proprietary and thus not available.

5.1 Normal Scenario in Example 1

The answer set of the program $\Pi(1)$ obtained according to our methodology contains the *occurs(a, i)* atoms shown in Appendix A, where **a** is an action and **i** is a time point on the reasoning time line. We use this case as a baseline when explaining the output of exceptional scenarios. As expected, the customer’s actions are interleaved with those of the waitress and cook. As well, an agent’s mental actions (i.e., selecting a goal and starting/ stopping an activity) take one time step when no other action of the same agent occur.

5.2 Serendipitous Achievement of Goal in Example 2

The logic form for this scenario is identical to the one for Example 1 shown in Example 5, except that the three observations about actions taking place at story time points 2–4 are replaced by

st_hpd(*pay*(*owner*, *b*), *true*, 2). *st_hpd*(*put_down*(*waitress*, *lentil_soup*, *t*), *true*, 3).
st_hpd(*eat*(*nicole*, *lentil_soup*), *true*, 4). *st_hpd*(*leave*(*nicole*), *true*, 5).

where *owner* is a new instance of sort *person*. Program $\Pi(2)$ has several answer sets, each one mapping the occurrence of action *pay*(*owner*, *b*) into a different reasoning time step from 12 to 18. Each of these answer sets also contains similar *occurs* atoms to $\Pi(1)$ up to time step 20 when Nicole eats the soup and the waitress stops her activity (given that the bill is already paid). From then on, it contains the following *occurs* predicates:

occurs(*stop*(*waitress*, *w_act*(*waitress*, *nicole*, *lentil_soup*, *lentil_soup*)), 20)
occurs(*eat*(*nicole*, *lentil_soup*), 20) *occurs*(*stand_up*(*nicole*), 23)
occurs(*start*(*nicole*, *c_subact_2*(*nicole*)), 21) *occurs*(*move*(*nicole*, *t*, *entrance*), 24)
occurs(*stop*(*nicole*, *c_subact_2*(*nicole*)), 22) *occurs*(*leave*(*nicole*), 25)
occurs(*stop*(*nicole*, *c_act*(*nicole*, *veg_r*, *lentil_soup*)), 26)

Thus, our system understands that Nicole has stopped *c_subact_2* immediately after starting it because she realized that its goal is already fulfilled. Note that approaches that do not view characters as goal-driven agents (including (Gabaldon 2009; Ng and Mooney 1992; Mueller 2007)) face substantial difficulties or simply cannot handle serendipitous scenarios. All answer sets of program $\Pi(2) \cup \{q1, q2\}$ contain the additional atoms

answer(*occur*(*leave*(*nicole*)), *yes*) *answer*(*occur*(*pay*(*nicole*, *b*)), *no*)

specifying that the answer to question *q1* is *Yes*, while the answer to *q2* is *No*.

5.3 Futile Activity in Example 3

The logic form for Example 3 contains the following observations:

st_hpd(*go*(*nicole*, *veg_r*), *true*, 0). *st_hpd*(*sit*(*nicole*), *true*, 1).
st_hpd(*pick_up*(*nicole*, *m*, *t*), *true*, 2). *st_obs*(*available*(*lentil_soup*, *veg_r*), *false*, 3).

and a fact *next_st*(2, 3) stating that story time steps 2 and 3 should be translated into consecutive reasoning time steps (see Section 4.3).

Program $\Pi(3)$ has multiple answer sets varying in the explanation for *st_obs*: exogenous action *make_unavailable*(*lentil_soup*, *veg_r*) occurs at a different reasoning time points in each of them. Otherwise, these answer sets contain the same *occurs* predicates as $\Pi(1)$ until step 9; then Nicole stops her futile activity and replans for her still active goal:

occurs(*stop*(*nicole*, *c_act*(*nicole*, *veg_r*, *waitress*, *lentil_soup*)), 10)
occurs(*replan*(*nicole*, *satiated_and_out*(*nicole*)), 11)

Thus the reader is cautious and does not make any assumptions about Nicole leaving the restaurant. As expected, it does not state that Nicole ate lentil soup either, which would be impossible, nor that she paid for anything.

5.4 Diagnosis in Example 4

The logic form for Example 4 contains a new instance of sort *food*, *miso_soup*, and the observations:

st_hpd(*go*(*nicole*, *veg_r*), *true*, 0).
st_hpd(*order*(*nicole*, *lentil_soup*, *waitress*), *true*, 1).
st_hpd(*put_down*(*waitress*, *miso_soup*, *t*), *true*, 2).

The program $\Pi(4)$ has two answer sets, containing explanations on what went wrong.

Answer Set 1. The first answer set indicates that the waitress started a different activity at time step 4 than the one in the normal scenario in Example 1:

`occurs(start(waitress, w_act(waitress, nicole, miso_soup, miso_soup)), 4)`

This activity can be read as “the waitress *understood* that Nicole ordered a miso soup and served her a miso soup.” Recall that a waiter’s possible activities are of the form $w_act(W, C, F_1, F_2)$, where F_1 may be a different food than the one actually ordered by the customer C , which allows reasoning about miscommunication between customer and waiter as in this answer set. Also, F_2 may be a different food than F_1 , meaning that the food served may not be the one that the waiter recorded as being ordered by the customer, which we will see in the second answer set. Our program considers all possible activities that can satisfy an agent’s goal and decides that the activity that can explain the later unfolding of the scenario must be the activity that the agent is actually executing.

This first answer set contains the same *occurs* predicates as $\Pi(1)$ up to time step 10, and then:

`occurs(interference, 11)`
`occurs(order(nicole, lentil_soup, waitress), 11)`
`occurs(move(waitress, t, kt), 12)`
`occurs(stop(nicole, c_subact_1(nicole, lentil_soup)), 12)`
`occurs(request(waitress, miso_soup, cook1), 13)`
`occurs(select(cook1, done_with_request(cook1, waitress)), 14)`
`occurs(start(cook1, ck_act(cook1, miso_soup, waitress)), 15)`
`occurs(prepare(cook1, miso_soup, waitress), 16)`
`occurs(stop(cook1, ck_act(cook1, miso_soup, waitress)), 17)`
`occurs(pick_up(waitress, miso_soup, kt), 17)`
`occurs(move(waitress, kt, t), 18)`
`occurs(put_down(waitress, miso_soup, t), 19)`

Here, the reader’s explanation is that there was some interference at time step 11 when Nicole ordered the lentil soup. As a result, the waitress misunderstood the order to be for miso soup and the cook followed her request.

Answer Set 2. The second answer set differs from the first in terms of what occurs at time steps 11-13, in particular

`occurs(interference, 13)`
`occurs(request(waitress, lentil_soup, cook1), 13)`

and the activities started by the waiter and cook at time steps 4 and 15:

`occurs(start(waitress, w_act(waitress, nicole, lentil_soup, miso_soup)), 4)`
`occurs(start(cook1, ck_act(cook1, miso_soup, waitress)), 15)`

It corresponds to a second possible explanation in which the waitress understood the order correctly and the misunderstanding/ interference occurred at step 13 when she communicated the order to the cook. (Note that the reader makes no assumptions about what happens next.)

Question Answering. These two answer sets show how reasoning by cases in ASP is useful to answering questions like: *Did the waitress ask the cook to prepare a lentil soup?* whose answer would be *No* in the first case (answer set 1) and *Yes* in the second one (answer set 2). In the future, we envision answering questions like *Why did Nicole receive a wrong order?* by producing the answers *A1: Because the waitress misunderstood the order./ A2: Because the cook misunderstood the order.*

5.5 Discussion

While we exemplified and tested our methodology on restaurant scenarios, our approach is equally applicable to other stereotypical activities, if we maintain the assumption that the knowledge base contains the relevant commonsense information. The main task when addressing a new stereotypical activity is defining the different \mathcal{TI} activities, including goals, of each involved agent. Part of this process can be automated by starting from a rigid and centralized script learned in an unsupervised manner (e.g., (Regneri et al. 2010)) and then assigning its actions to the activities of different agents, those performing them. Determining (sub-)goals and splitting activities into sub-activities is a more challenging problem, which deserves substantial attention.

6 Conclusions and Future Work

In this work, we proposed a new methodology for automating the understanding of narratives about stereotypical activities. As a first main contribution, we overcame limitations of Mueller’s work (2007) related to exceptional scenarios. To achieve this, we had to abandon Mueller’s rigid script-based approach and use instead a substantial portion of a state-of-the-art intelligent agent architecture, augmented with support for intentions. Our second contribution is extending the architecture of an intentional agent, \mathcal{AIA} (Blount et al. 2015), to model a third-person observer. We exemplified our methodology on several types of scenarios. In the future, we intend to create an extensive story corpus, which will be a laborious task, to further test our methodology.

References

- BALDUCCINI, M. 2007. CR-MODELS: An inference engine for CR-Prolog. In *Proceedings of LPNMR 2007*, C. Baral, G. Brewka, and J. S. Schlipf, Eds. LNCS, vol. 4483. Springer, 18–30.
- BALDUCCINI, M., BARAL, C., AND LIERLER, Y. 2007. *Handbook of Knowledge Representation. Foundations of Artificial Intelligence*. Elsevier, Chapter 20. Knowledge Representation and Question Answering.
- BALDUCCINI, M. AND GELFOND, M. 2003. Logic Programs with Consistency-Restoring Rules. In *Proceedings of Commonsense-03*. AAAI Press, 9–18.
- BALDUCCINI, M. AND GELFOND, M. 2008. The AAA architecture: An overview. In *Architectures for Intelligent Theory-Based Agents, Papers from the 2008 AAAI Spring Symposium, 2008*. AAAI Press, 1–6.
- BARAL, C. AND GELFOND, M. 2005. Reasoning about Intended Actions. In *Proceedings of AAAI-05*. AAAI Press, 689–694.
- BLOUNT, J. 2013. An Architecture for Intentional Agents. Ph.D. thesis, Texas Tech University, Lubbock, TX, USA.
- BLOUNT, J., GELFOND, M., AND BALDUCCINI, M. 2015. A theory of intentions for intelligent agents. In *Proceedings of LPNMR 2015*, F. Calimeri, G. Ianni, and M. Truszczyński, Eds. LNCS, vol. 9345. Springer, 134–142.
- BORDINI, R. H., HÜBNER, J. F., AND WOOLDRIDGE, M. 2007. *Programming Multi-Agent Systems in AgentSpeak Using Jason*. John Wiley & Sons, Ltd.
- DIAKIDOY, I.-A., KAKAS, A., MICHAEL, L., AND MILLER, R. 2015. Star: A system of argumentation for story comprehension and beyond. 12th International Symposium on Logical Formalizations of Commonsense Reasoning (Commonsense’15). 64–70.

- GABALDON, A. 2009. Activity recognition with intended actions. In *Proceedings of IJCAI 2009*, C. Boutilier, Ed. 1696–1701.
- GELFOND, M. AND KAHL, Y. 2014. *Knowledge Representation, Reasoning, and the Design of Intelligent Agents*. Cambridge University Press.
- GELFOND, M. AND LIFSCHITZ, V. 1991. Classical Negation in Logic Programs and Disjunctive Databases. *New Generation Computing* 9, 3/4, 365–386.
- INCLEZAN, D., ZHANG, Q., BALDUCCINI, M., AND ISRANEY, A. 2017. Understanding restaurant stories using an ASP theory of intentions (Extended abstract). In *Technical Communications of the 33rd International Conference on Logic Programming (ICLP-TC 2017)*. OASICs.
- KAMP, H. AND REYLE, U. 1993. *From discourse to logic*. Vol. 1,2. Kluwer.
- LIERLER, Y., INCLEZAN, D., AND GELFOND, M. 2017. Action languages and question answering. In *IWCS 2017 - 12th International Conference on Computational Semantics - Short papers*.
- MICHAEL, L. 2013. Story understanding... calculemus! 11th International Symposium on Logical Formalizations of Commonsense Reasoning (Commonsense'13). 64–70.
- MOSTAFAZADEH, N., VANDERWENDE, L., YIH, W.-T., KOHLI, P., AND ALLEN, J. 2016. Story cloze evaluator: Vector space representation evaluation by predicting what happens next. In *Proceedings of RepEval'16*. Association for Computational Linguistics, 24–29.
- MUELLER, E. T. 2004. Understanding script-based stories using commonsense reasoning. *Cognitive Systems Research* 5, 4, 307–340.
- MUELLER, E. T. 2007. Modelling space and time in narratives about restaurants. *Literary and Linguistic Computing* 22, 1, 67–84.
- NG, H. T. AND MOONEY, R. J. 1992. Abductive plan recognition and diagnosis: A comprehensive empirical evaluation. In *Proceedings of the 3rd International Conference on Principles of Knowledge Representation and Reasoning (KR'92)*. 499–508.
- NIEVES, J. C., GUERRERO, E., AND LINDGREN, H. 2013. Reasoning about human activities: an argumentative approach. In *Twelfth Scandinavian Conference on Artificial Intelligence, SCAI 2013, Aalborg, Denmark, November 20-22, 2013*. 195–204.
- PALMER, M., GILDEA, D., AND KINGSBURY, P. 2005. The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics* 31, 1 (Mar.), 71–106.
- RAO, A. S. AND GEORGEFF, M. P. 1991. Modeling rational agents within a BDI-architecture. In *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning (KR'91)*. Cambridge, MA, USA, April 22-25, 1991. 473–484.
- REGNERI, M., KOLLER, A., AND PINKAL, M. 2010. Learning script knowledge with web experiments. In *Proceedings of ACL '10*. 979–988.
- RICHARDSON, M., BURGESS, C. J. C., AND RENSHAW, E. 2013. Mctest: A challenge dataset for the open-domain machine comprehension of text. In *EMNLP. ACL*, 193–203.
- SCHANK, R. C. AND ABELSON, R. P. 1977. *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures*. Lawrence Erlbaum.
- SHANAHAN, M. 1997. *Solving the Frame Problem*. MIT Press.
- TODOROVA, Y. AND GELFOND, M. 2012. Toward Question Answering in Travel Domains. In *Correct Reasoning*. 311–326.
- WOOLDRIDGE, M. 2009. *An Introduction to MultiAgent Systems*, 2nd ed. Wiley Publishing.
- ZHANG, Q. AND INCLEZAN, D. 2017. An application of ASP theories of intentions to understanding restaurant scenarios. In *Proceedings of PAoASP'17*.

Appendix A Partial Output for the Normal Scenario in Example 1

```

map(0, 2)
map(1, 11)
map(2, 19)
map(3, 20)
map(4, 31)
occurs(select(nicole, satiated_and_out(nicole)), 0)
occurs(start(nicole, c_act(nicole, veg_r, waitress, lentil_soup)), 1)
occurs(go(nicole, veg_r), 2)
occurs(select(waitress, served_and_billed(nicole)), 3)
occurs(start(waitress, w_act(waitress, nicole, lentil_soup, lentil_soup)), 4)
occurs(greet(waitress, nicole), 5)
occurs(lead_to(waitress, nicole, t), 6)
occurs(sit(nicole), 7)
occurs(start(nicole, c_subact_1(nicole, lentil_soup, waitress)), 8)
occurs(pick_up(nicole, m, t), 9)
occurs(put_down(nicole, m, t), 10)
occurs(order(nicole, lentil_soup, waitress), 11)
occurs(stop(nicole, c_subact_1(nicole, lentil_soup, waitress)), 12)
occurs(move(waitress, t, kt), 12)
occurs(request(waitress, lentil_soup, cook1), 13)
occurs(select(cook1, done_with_request(cook1, waitress)), 14)
occurs(start(cook1, ck_act(cook1, lentil_soup, waitress)), 15)
occurs(prepare(cook1, lentil_soup, waitress), 16)
occurs(stop(cook1, ck_act(cook1, lentil_soup, waitress)), 17)
occurs(pick_up(waitress, lentil_soup, kt), 17)
occurs(move(waitress, kt, t), 18)
occurs(put_down(waitress, lentil_soup, t), 19)
occurs(eat(nicole, lentil_soup), 20)
occurs(start(nicole, c_subact_2(nicole, waitress)), 21)
occurs(request(nicole, b, waitress), 22)
occurs(move(waitress, t, ct), 23)
occurs(pick_up(waitress, b, ct), 24)
occurs(move(waitress, ct, t), 25)
occurs(put_down(waitress, b, t), 26)
occurs(stop(waitress, w_act(waitress, nicole, lentil_soup, lentil_soup)), 27)
occurs(pay(nicole, b), 27)
occurs(stop(nicole, c_subact_2(nicole, waitress)), 28)
occurs(stand_up(nicole), 29)
occurs(move(nicole, t, entrance), 30)
occurs(leave(nicole), 31)
occurs(stop(nicole, c_act(nicole, veg_r, waitress, lentil_soup)), 32)

```