

Story Comprehension through Argumentation¹

Irene-Anna DIAKIDOY ^a, and Antonis KAKAS ^b and Loizos MICHAEL ^c and
Rob MILLER ^d

^a *University of Cyprus, eddiak@ucy.ac.cy*

^b *University of Cyprus, antonis@cs.ucy.ac.cy*

^c *Open University of Cyprus, loizos@ouc.ac.cy*

^d *University College London, rsm@ucl.ac.uk*

Abstract. This paper presents a novel application of argumentation for automated Story Comprehension (SC). It uses argumentation to develop a computational approach for SC as this is understood and studied in psychology. Argumentation provides uniform solutions to various representational and reasoning problems required for SC such as the frame, ramification, and qualification problems, as well as the problem of contrapositive reasoning with default information. The grounded semantics of argumentation provides a suitable basis for the construction and revision of comprehension models, through the synthesis of the explicit information from the narrative in the text with the implicit (in the reader's mind) common sense world knowledge pertaining to the topic(s) of the story given in the text. We report on the empirical evaluation of the approach through a prototype system and its ability to capture both the majority and the variability of understanding of stories by human readers. This application of argumentation can provide an important test-bed for the more general development of computational argumentation.

1. Introduction

Argumentation is prevalent in many forms of human reasoning. Recently, new psychological evidence [13] has re-enforced the close link between argumentation and human reasoning suggesting that this is, in its general form, inherently argumentative. It is, therefore, important to connect developments of computational argumentation and its theory to applications related to human reasoning. Our broader application aim is to develop cognitive systems for semantically analyzing information with a narrative structure; e.g., news feeds over the Web or dialogues over social media. Our broader scientific aim is to ascertain the extent to which concepts and theories from the field of human psychology, and in particular the psychology of human story comprehension, can inform research into automated cognitive systems. Our working hypothesis is that such psychological concepts are useful and indeed necessary for both building and testing such automated systems, and the present paper reports on one stage in testing that hypothesis.

This paper presents a novel application of argumentation to a particular case of text comprehension, that of *story comprehension* (SC). In particular, the *evaluation method-*

¹A short high-level summary (without any technical details) of part of this work appears in KR 2014.

ology that we use is: (i) set up a corpus of stories with questions to test different aspects of story comprehension; (ii) harness the world knowledge used by human readers for comprehension; (iii) use this world knowledge in our framework and automated system, and compare its comprehension behaviour with that of the human readers. Given the strong link of our work with psychology, it is useful to give here a brief summary of the problem and central notions of SC as identified by research in psychology.

1.1. Psychological Background

Comprehending narrative texts entails the construction of a mental representation of the information contained in the text. Since, no narrative specifies clearly and completely all implications or the relations between them, comprehension depends also on the ability to generate *bridging and elaborative inferences* that connect and elaborate it resulting in a mental or *comprehension model* of the narrative. Inference generation is necessary to comprehend any narrative text as a whole, i.e., as a single network of interconnected propositions instead of as a series of isolated sentences, and to appreciate the suspense and surprise that characterize narrative texts or stories, in particular [4,12].

Although inference generation is based on the activation of background *world knowledge*, the process is constrained by text information. Concepts encountered in the text activate related knowledge in the readers' long-term memory [9]. Nevertheless, at any given point in the process, only a small subset of all the possible knowledge-based inferences remain activated and become part of the mental representation: those that connect and elaborate text information in a way that contributes to the *coherence* of the mental model [12,25]. Inference generation is a task-oriented process that follows the principle of *cognitive economy* enforced by a limited-resource cognitive system.

Since the results of this coherence-driven selection mechanism can easily exceed the limited working memory capacity of the human cognitive system, coherence on a more global level is achieved through *higher-level integration* processes that create macro-propositions generalizing or subsuming a number of text-encountered concepts and the inferences that connect them. Previously selected information with few connections to other parts is dropped from the mental model, resulting in a more consolidated network of propositions, which serves as the new anchor for processing subsequent text [10].

Comprehension also requires an iterative *revision mechanism* of the readers' mental model. The feelings of suspense and surprise that stories aim to create are achieved through discontinuities or changes (in settings, motivations, actions, or consequences) that are not predictable or are wrongly predictable solely on the basis of the mental model created so far. Knowledge about the structure and the function of stories leads readers to expect discontinuities and to use them as triggers to revise their mental model [29]. A change in time or setting in the text may serve as a clue for revising parts of the mental model while other parts remain and are integrated with subsequent text information.

Finally, the interaction of processes for coherence carries the possibility of *different* but equally *successful comprehension* outcomes, due to the qualitative and quantitative differences in the conceptual and mental state knowledge of different readers.

1.2. Approach and Scope of the Present Paper

Our approach will be based on developing a preference-based argumentation framework for SC, using standard argumentation semantics, such as that of the grounded extension,

to formalize the notion of a comprehension model. This framework will uniformly encompass a Reasoning about Actions and Change (RAC) framework for the temporal development of the information in a story together with Default Reasoning with the relevant parts of the world knowledge pertaining to the story. Notions from preference-based argumentation (e.g., [20]) are used for building arguments and attacks between them. We also use ideas from [6,8] to allow us to employ contrapositive reasoning, despite the defeasible nature of the available knowledge, in building arguments from a given story.

At present we concentrate on representing narratives and the *world knowledge* needed for the central comprehension process of *synthesizing and elaborating* the explicit text information with new inferences, and *revising* them in the presence of new narrative information. Our working hypothesis is that higher level features of comprehension, such as coherence and cognitive economy, can be tackled on top of the framework we develop. We are also assuming as solved the issue of correctly parsing the natural language of the text into some information-equivalent structured (e.g., logical) form of the story narrative, without discounting the importance of this problem, nor the possibility of the need to be tackled in conjunction with the problems on which we are focusing.

We will use a story from the initial evaluation of our approach as a running example:

It was the night of Christmas Eve. After feeding the animals and cleaning the barn, Papa Joe took his shotgun from above the fireplace and sat out on the porch cleaning it. He had had this shotgun since he was young, and it had never failed him, always making a loud noise when it fired.

Papa Joe woke up early at dawn, picked up his shotgun and went off to the forest. He walked for hours, until the sight of two turkeys in the distance made him stop suddenly. A bird on a tree nearby was cheerfully chirping away, building its nest. He aimed at the first turkey, and pulled the trigger.

After a moment's thought, he opened his shotgun and saw there were no bullets in the shotgun's chamber. He loaded his shotgun, aimed at the turkey and pulled the trigger again. Undisturbed, the bird nearby continued to chirp and build its nest. Papa Joe was very confused. Would this be the first time that his shotgun had let him down?

Section 2 defines the argumentation-based semantics for SC. Section 3 presents the empirical evaluation of the approach. The paper concludes with related and future work.

2. Argumentation Framework

The construction of a comprehension model and its qualification and revision at all levels as the story unfolds, is captured through a uniform acceptability requirement on the arguments that support the conclusions in the model. We use methods and results from Argumentation Theory in AI [5,20], and link these to Reasoning about Action and Change (RAC) with Default Reasoning on the static properties of the domains of discourse.

2.1. Story Representation

We start by defining a **story representation** as a triple, $\mathcal{SR} = \langle \mathcal{N}, \mathcal{W}, \succ \rangle$, comprising the narrative \mathcal{N} , the world knowledge \mathcal{W} used for comprehension, and a priority relation

\succ . For the representation we use a typical RAC language of Fluents, Actions, and Times. The exact time-points are largely inconsequential, and stand for the abstract scenes in a story. Narratives are represented as a sequence of observations, stating what holds / occurs and when, as presented in the story. In our example story ($pj = \text{“Papa Joe”}$):

OBS(*alive*(turkey), 1), OBS(*aim*(pj, turkey), 1), OBS(*pull_trigger*(pj), 1),
 OBS(\neg *gun_loaded*, 4), OBS(*load_gun*, 5), OBS(*pull_trigger*(pj), 6),
 OBS(*chirp*(bird), 10), OBS(*nearby*(bird), 10).

World knowledge is represented as a collection of **unit-arguments** of the form $\text{arg}(H, B)$, where H is a fluent / action literal, and B is a set of such. Each unit-argument captures a simple association between concepts in the language: if the body B holds, then we have some evidence to believe that the head H holds. This stems from a key observation in psychology that typically all world knowledge, and irrespective of type, is inherently default and is not fully qualified at the representation level. It is qualified via the reasoning process by the relative strength of other (conflicting) knowledge. There are four different types of unit-arguments capturing different types of world knowledge: **causal unit-arguments** $\text{cau}(H, B)$ capture how properties are caused to come about; **property unit-arguments** $\text{pro}(H, B)$ capture how properties relate to each other; **preclusion unit-arguments** $\text{prc}(H, B)$ capture how properties preclude other properties for changing; **persistence unit-arguments** $\text{per}(H, \{H\})$ capture how properties persist over time. Persistence unit-arguments need not be explicitly represented in the world knowledge \mathcal{W} , and are implicitly assumed to be present for each literal H in the RAC language. We sometimes write $\text{per}(H)$ to mean $\text{per}(H, \{H\})$, as a way to improve readability.

The priority relation \succ in a story representation is defined on unit-arguments. In general, this includes (i) $\text{prc}(H, B_1) \succ \text{cau}(\neg H, B_2)$; (ii) $\text{cau}(H, B_1) \succ \text{per}(\neg H, B_2)$; (iii) $\text{per}(H, B_1) \succ \text{pro}(\neg H, B_2)$; and (iv) story-specific or knowledge-specific priorities between unit-arguments in \mathcal{W} .

For our example story one could consider the following world knowledge:

$c1 : \text{cau}(\textit{fired_at}(pj, X), \{\textit{aim}(pj, X), \textit{pull_trigger}(pj)\})$
 $c2 : \text{cau}(\neg \textit{alive}(X), \{\textit{fired_at}(pj, X), \textit{alive}(X)\})$
 $c3 : \text{cau}(\textit{noise}, \{\textit{fired_at}(pj, X)\})$
 $c4 : \text{cau}(\neg \textit{chirp}(bird), \{\textit{noise}, \textit{nearby}(bird)\})$
 $c5 : \text{cau}(\textit{gun_loaded}, \{\textit{load_gun}\})$
 $r1 : \text{prc}(\neg \textit{fired_at}(pj, X), \{\neg \textit{gun_loaded}\})$
 $p2 : \text{pro}(\neg \textit{fired_at}(pj, X), \{\neg \textit{noise}\})$

with the extra story-specific and knowledge-specific priorities $r1 \succ c1$, and $p2 \succ c1$.

The use of priorities between unit-arguments addresses the *endogenous qualification problem*, while the priority of information in the narrative over any unit-argument (formalized in the sequel) also addresses the *exogenous qualification problem*. In addition, the last two general priorities given above address the (*generalized*) *frame problem*, ensuring that properties cease to persist when and only when there is causal evidence to the contrary, even in the case where property laws remain violated by this persistence.

2.2. Drawing Inferences — Constructing Arguments

To account for the use of unit-arguments to draw inferences in the temporal setting of a story, we introduce the notion of an **argument-rule** $\text{arg}(H, B)@T^h \xrightarrow{d} (C, T)$, com-

prising a unit-argument $\text{arg}(H, B)$, the time-point T^h at which the head of the unit-argument head is applied, and the conclusion (C, T) that follows from its application, where C is a fluent / action literal, and T the time-point at which the literal is inferred to hold. The body of a unit-argument is applied at time-point T^b which equals T^h for property unit-arguments, and equals $T^h - 1$ for all other unit-arguments. An argument-rule can use its unit-argument $\text{arg}(H, B)$ in the usual forward direction $d = \text{F}$ or in a backward direction $d = \text{B}$. In the former case, the **conclusion** (C, T) is (H, T^h) and the **premise** $\{(L, T^b) \mid L \in B\}$, whereas in the latter case, the **conclusion** (C, T) is $(\neg X, T^b)$ for some $X \in B$ and the **premise** is $\{(L, T^b) \mid L \in B \setminus \{X\}\} \cup \{(\neg H, T^h)\}$.

The use of argument-rules in the backward direction allows the framework to include *reasoning by contradiction*, e.g., contraposition, with the defeasible nature of the world knowledge. It gives for example a form of *backward persistence*, e.g., from an observation to support (but not necessarily conclude, due to a possible qualification by other pieces of knowledge) that the observed property holds also at previous time-points.

Definition 1. A *timed literal* (C, T) is a **supported conclusion** of a set A of argument-rules if $\text{OBS}(C, T) \in \mathcal{N}$, or if (C, T) is the conclusion of an argument-rule in A . A set of argument-rules A is **story-grounded** if it can be totally ordered so that every (L, T) in the premise of any argument-rule in A is a supported conclusion of the set of argument-rules that precede the aforementioned argument-rule in the chosen ordering of A .

2.3. Argumentation Semantics

Given a story representation $\mathcal{SR} = \langle \mathcal{N}, \mathcal{W}, \succ \rangle$, we define a corresponding abstract argumentation framework $\langle \mathcal{A}^{\mathcal{SR}}, \text{Att}^{\mathcal{SR}} \rangle$ following the two key suggestions from psychology: inferences drawn by readers are (i) grounded on the explicit information in the story narrative, and (ii) sceptical in nature. The first suggestion leads to the next definition:

Definition 2 (Arguments). An **argument** in $\mathcal{A}^{\mathcal{SR}}$ is any story-grounded set of argument-rules. (C, T) is an **inference** of A if it is a supported conclusion of A .

In defining the attacking relation, $\text{Att}^{\mathcal{SR}}$, we need to consider carefully the subtleties of backward reasoning through the defeasible unit-arguments. In general, for arguments to attack each other they need to be in conflict, e.g., draw conflicting conclusions. The use of contrapositive reasoning for backward inference also means that it is possible to have attacking arguments that support conclusions that are not in direct conflict, but instead whose unit-arguments have conflicting heads. For instance, in our running example we can use the causal unit-argument $c1$ in A_1 to forward derive $\text{fired_at}(pj, X)$ and the preclusion unit-argument $r1$ in A_2 to backward derive gun_loaded from $\text{fired_at}(pj, X)$; despite the fact that the derived conclusions of A_1 and A_2 are not in conflict, the unit-arguments used have conflicting heads. Although not all such *indirect* conflicts are important, a certain subset does need to be accounted for.

Definition 3 (Conflicts). Consider two argument-rules $\rho_1 = \text{arg}_1(H_1, B_1) @ T_1^h \xrightarrow{d_1} (C_1, T_1)$ and $\rho_2 = \text{arg}_2(H_2, B_2) @ T_2^h \xrightarrow{d_2} (C_2, T_2)$. These argument-rules are **in direct conflict** if $C_1 = \neg C_2$, $T_1 = T_2$; they are **in indirect conflict** if $H_1 = \neg H_2$, $T_1^h = T_2^h$.

Informally, an argument will attack another if the former includes an argument-rule that is in conflict with an argument-rule in the latter and the attacking argument-rule is not weaker in terms of the priority relation on their respective unit-arguments. When the conflict is indirect care needs to be taken when reasoning backwards with an argument-rule and the support of the head of this argument-rule comes from a stronger argument-rule used in the forward direction. For example, consider the two usual unit-arguments about birds, penguins, and their (in)ability to fly, along with their preference $\text{pro}(\neg\text{fly}, \{\text{penguin}\}) \succ \text{pro}(\text{fly}, \{\text{bird}\})$. Given the observation $\text{OBS}(\text{penguin}, 1)$, one may apply the first unit-argument in the forward direction to derive $(\neg\text{fly}, 1)$. However, it is not permissible to subsequently apply the second unit-argument in the backward direction to derive $(\neg\text{bird}, 1)$ and an attack will exist to prevent this. On the other hand, given the observation $\text{OBS}(\text{bird}, 1)$, one may apply the second unit-argument in the forward direction to derive $(\text{fly}, 1)$, and subsequently apply the first unit-argument in the backward direction to derive $(\neg\text{penguin}, 1)$. This distinction is reflected in the next two definitions. The full treatment of this is beyond the scope of this paper; it would involve an extended semantics for argumentation, such as that for Argumentation Logic [6,8], where proof by contradiction is reconstructed in terms of argumentation.

Definition 4 (Qualification). Consider two argument-rules $\rho_1 = \text{arg}_1(H_1, B_1) @ T_1^h \xrightarrow{d_1} (C_1, T_1)$ and $\rho_2 = \text{arg}_2(H_2, B_2) @ T_2^h \xrightarrow{d_2} (C_2, T_2)$. Argument-rule ρ_1 (**endogenously**) **qualifies** argument-rule ρ_2 if $\text{arg}_2(H_2, B_2) \not\prec \text{arg}_1(H_1, B_1)$, and either ρ_1 and ρ_2 are in direct conflict, or they are in indirect conflict and $d_2 = \text{F}$, $d_1 = \text{B}$. In particular, if $\text{arg}_1(H_1, B_1) \succ \text{arg}_2(H_2, B_2)$, then ρ_1 **strongly qualifies** ρ_2 ; otherwise, ρ_1 **weakly qualifies** ρ_2 . The **story (exogenously) qualifies** argument-rule ρ_2 if $\text{OBS}(\neg C, T_2) \in \mathcal{N}$.

Definition 5 (Attacking Relation). An argument A_1 **attacks** an argument A_2 , and thus $(A_1, A_2) \in \text{Att}^{SR}$, if an argument-rule ρ_1 in A_1 strongly qualifies an argument-rule ρ_2 in A_2 , or ρ_1 weakly qualifies ρ_2 and there is no argument-rule ρ'_1 in A_1 that is strongly qualified by an argument-rule ρ'_2 in A_2 . Furthermore, the empty argument **attacks** an argument A_2 , and thus $(\emptyset, A_2) \in \text{Att}^{SR}$, if the story qualifies an argument-rule in A_2 .

The definition of attack anticipates their use in the definition of a comprehension model, where it is the *minimal* attacking arguments that can render some other argument not suitable. In such minimal attacks all argument-rules, ρ'_1 , that iteratively support the premise of ρ_1 must not be strongly qualified by some argument-rule ρ'_2 in A_2 . Following the guideline from psychology for sceptical inferences, we can select the grounded extension semantics to define the central notion of comprehension model.

Definition 6 (Comprehension Model). Given a story SR and the corresponding argumentation framework $\langle \mathcal{A}^{SR}, \text{Att}^{SR} \rangle$, a set of arguments $\Delta \subseteq \mathcal{A}^{SR}$ is a **comprehension model** of SR if Δ is a subset of the (unique) grounded extension of $\langle \mathcal{A}^{SR}, \text{Att}^{SR} \rangle$.

As suggested by psychology, not all possible (sceptical) inferences are, or should be, drawn when reading a story and hence any subset of the grounded extension can be used. These subsets need not contain all their defending arguments. Comprehension models that contain explicitly their defenses are also required to be *admissible*.

A comprehension model can be tested, as is often done in psychology, through a series of multiple-choice questions with answers of the form “ C holds at T ”.

Definition 7. Let M be a comprehension model of a story representation SR . An answer of the form “ C holds at T ” is **accepted** if (C, T) is a supported conclusion of M , it is **rejected** if $(\neg C, T)$ is a supported conclusion of M , and it is **possible** otherwise.

2.4. Reasoning Illustration

To illustrate the formal framework, how arguments are constructed and how a comprehension of a story is formed, let us consider our example story starting from the end of the second paragraph corresponding to time-points 1–3 in the example narrative. Note that the empty argument A_1 supports $(aim(pj, turkey), 1)$ and $(pull_trigger(pj), 1)$. Hence, $c1$ on 2 forward concludes $(fired_at(pj, turkey), 2)$ under the empty argument A_1 . We can thus populate A_1 with $c1 @ 2 \xrightarrow{F} (fired_at(pj, turkey), 2)$. Similarly, we can include $per(alive(turkey)) @ 2 \xrightarrow{F} (alive(turkey), 2)$ in the new A_1 . Under this latter A_1 , $c2$ on 3 forward concludes $(\neg alive(turkey), 3)$, allowing us to further extend A_1 with $c2 @ 3 \xrightarrow{F} (\neg alive(turkey), 3)$. The resulting A_1 is an argument that supports $(\neg alive(turkey), 3)$. It is based on this inference, that we expect readers to respond that the first turkey is dead, when asked about its status at this point, since no other argument grounded on the narrative (thus far) can support a qualification to this inference (and hence attack). Note, also, that we can further include in A_1 the argument-rule $r1 @ 2 \xrightarrow{B} (gun_loaded, 1)$ to support, using backward (contrapositive) reasoning with $r1$, the conclusion that the gun was loaded when its trigger was pulled at time-point 1.

Reading the first sentence of the third paragraph, we learn $OBS(\neg gun_loaded, 4)$. We expect that this new piece of evidence will lead readers to revise their inferences, as now we have an argument that supports the conclusion $(\neg fired_at(pj, turkey), 2)$ based on the stronger (qualifying) unit-argument $r1$. For this we need to support the premise $\{(\neg gun_loaded, 1)\}$ of the argument-rule $r1 @ 2 \xrightarrow{F} (\neg fired_at(pj, turkey), 2)$. We can do so by using the three argument-rules $per(gun_loaded) @ 4 \xrightarrow{B} (\neg gun_loaded, 3)$, $per(gun_loaded) @ 3 \xrightarrow{B} (\neg gun_loaded, 2)$, $per(gun_loaded) @ 2 \xrightarrow{B} (\neg gun_loaded, 1)$, which support the conclusion that the gun was also unloaded before it was observed to be so. This uses $per(gun_loaded)$ contrapositively, effectively reasoning through a proof by contradiction: had the gun been loaded at 1, it would have been so also at 2, 3, and 4, which would contradict the story. Note, though, that this backward inference of $\neg gun_loaded$ would be qualified if the world knowledge contained the unit-argument $c : cau(\neg gun_loaded, \{pull_trigger(pj)\})$. This latter unit-argument would lead to an indirect conflict at time-point 2 with the backward persistence of $\neg gun_loaded$ from 2 to 1 and due to the stronger nature of causal over persistence unit-arguments the argument-rule corresponding to the backward persistence of $\neg gun_loaded$ would be qualified.

Assuming that c is absent, the argument A_2 consisting of the three “persistence” argument-rules is in conflict on $(gun_loaded, 1)$ with the argument A_1 above. Each argument attacks the other, and neither can be part of a comprehension model. If we extend A_2 with $r1 @ 2 \xrightarrow{F} (\neg fired_at(pj, turkey), 2)$ then this can now attack A_1 using the priority of $r1$ over $c1$. The weak qualification of the backward “persistence” argument-rules in A_2 by $r1 @ 2 \xrightarrow{B} (gun_loaded, 1)$ in A_1 no longer leads to an attack from A_1 to A_2 , since $r1 @ 2 \xrightarrow{F} (\neg fired_at(pj, turkey), 2)$ strongly qualifies an argument-rule in A_1 .

Therefore, the extended A_2 is part of a comprehension model and the conclusion $(\neg fired_at(pj, turkey), 2)$ is drawn revising the previous conclusions drawn from

A_1 . The process of understanding our story may then proceed by extending A_2 with $\text{per}(alive(turkey))@ T \xrightarrow{F} (alive(turkey), T)$ for $T = 2, 3, 4$, resulting in a comprehension model that infers $alive(turkey)$ at 4. It is based on this inference that we expect readers to respond that the first turkey is alive at 4 after reading the story so far.

Continuing with the story, after Papa Joe loads the gun and fires again, we can support by forward inferences that the gun fired, that noise was caused, and that the bird stopped chirping, through a chaining of the causal unit-arguments $c1, c3, c4$. But $\text{OBS}(chirp(bird), 10)$ allows the construction of arguments that attack on all these through the repeated backward use of the same unit-arguments grounded on this observation. We thus have an exogenous qualification effect where these conclusions can not be sceptical and so will not be inferred by any comprehension model. But if we also consider the stronger information in $p2$, that this gun does not fire without a noise, together with the backward conclusion of $\neg noise$, an argument that contains these can attack the firing of the gun at time-point 2 and thus defend against attacks that are grounded on $\text{OBS}(pull_trigger(pj), 1)$ and the gun firing. As a result, we have the effect of blocking the ramification of the causation of $noise$ and so $\neg noise$ (and $\neg fired_at(pj, turkey)$) are sceptically concluded. Readers, indeed respond in this way at this point in the story.

With this latter part of the example story we see how our framework addresses the *ramification problem* and its non-trivial interaction with the qualification problem [27]. In fact, a generalized form of this problem is addressed where the ramifications are not chained only through causal laws but through any of the forms of inference we have in the framework — causal, property, preclusion, or persistence — and through any of the types of inference — forward or backward by contradiction. Weak links in this chain of ramifications that happen to be qualified, effectively break the chain of inferences that would otherwise be supported. Note, also, that the intermediate ramifications might be realized over a sequence of time-points, which in the context of this work are better thought of as micro/inference-level time-points that are more dense than the macro/story-level time-points; a more explicit distinction of micro/macro time-points is taken in [7].

3. Framework Evaluation

We have proceeded to evaluate our argumentation-based approach to SC following the three step general evaluation methodology described in the introduction. To this end, we have implemented an argumentation system able to read a story narrative in a sequence of sessions, compute a comprehension model (the maximal one) based on the read parts of the story, answer (multiple choice) questions, and repeat the process, revising its conclusions as more parts of the story narrative become available. The task is to evaluate the inferences of the system against those of human readers when using world knowledge as that used by human readers that was obtained through empirical psychological studies.

The efficiency of the system comes from the known polynomial-time computational complexity of computing the grounded extension of an argumentation framework, but also by the direct manipulation of argument-rules (rather than arguments) when computing a comprehension model. To further mitigate the possibly exponential number of arguments in the size of the background knowledge, one would need to apply *coherency* and *cognitive economy* operations on top of these underlying argumentation processes. Such operations can again be guided from psychological studies, but this is beyond the

scope of this paper. The initial development of the system is meant as a proof of principle to be applied on relatively short stories and small vocabulary of world knowledge.

3.1. System Implementation

The system has been implemented in Prolog, along with an accompanying high-level language for representing narratives, background knowledge, and multiple-choice questions. The system is available at <http://cognition.ouc.ac.cy/narrative/>.

Without going into details, the language allows the user to specify a sequence of sessions of the form `session(s(B), Qs, Vs)`, specifying the scene $s(B)$ of the story to read, the questions Qs to answer, and the parts Vs of the comprehension model to be made visible to the user. The narrative is represented by a sequence of statements of the form $s(B) :: X \text{ at } T$. The background knowledge is represented by clauses of the form $p(N) :: A, B, \dots, C \text{ implies } X$, or $r(N) :: A, B, \dots, C \text{ precludes } X$, or $c(N) :: A, B, \dots, C \text{ causes } X$, where p, r, c shows, respectively, a property, preclusion, or causal unit-argument, named N . Negations are represented by prefixing a fluent or action with the minus symbol. Variables are used to represent relational information. Preferences between unit-arguments are represented in the form $p(N1) \gg c(N2)$. Questions are represented by clauses of the form $q(N) ?? (X1 \text{ at } T1, \dots, X2 \text{ at } T2) ; \dots$, where N is the question name, $(X1 \text{ at } T1, \dots, X2 \text{ at } T2)$ is the first possible answer as a conjunction of fluents or actions that need to hold at their respective time-points, and $;$ separates the answers.

The implemented system demonstrates real modularity and elaboration tolerance, allowing as input any story narrative or background knowledge in the given syntax, appropriately qualifying the given information to compute a comprehension model.

3.2. Empirical Evaluation

Following our evaluation methodology, we have carried a *psychological study* to ascertain the world knowledge that is activated to successfully comprehend stories on the basis of data obtained from human readers. We developed a set of inferential questions that were presented to participants after reading pre-specified story segments. The experimental materials are all available at <http://cognition.ouc.ac.cy/narrative/>. Stories and questions were presented to the participants in their natural language form, and assessed the extent to which readers connected, explained, and elaborated key story elements. Readers were instructed to answer each question and to justify their answers using a “think-aloud” method of answering questions while reading, in order to reveal the world knowledge they had used. The readers did not interact with the automated system, nor did they have access to the formal representations of the stories and questions.

The qualitative data from the readers was pooled together and analysed as to the frequencies of the types of responses in conjunction with the information given in justifications and think-aloud protocols. The gathered data was formally represented, and was used both to provide the automated system with the background knowledge needed to draw inferences, and also to provide a yardstick against which to evaluate the automated system’s performance. Considering those readers that demonstrated successful comprehension according to psychological criteria, our automated system was able to identify the most popular answer to each question, and also to recognize questions for which no single answer was accepted, following the variability demonstrated by human readers.

To illustrate this let us consider our example story and two questions: “01: Where did Papa Joe live?” and “06: What was Papa Joe doing in the woods?”. The parts of the story representation relevant to these are:

s(1) :: night at 0.	s(2) :: animal(turkey2) at 2.
s(1) :: xmasEve at 0.	s(2) :: alive(turkey1) at 2.
s(1) :: clean(pj,barn) at 0.	s(2) :: alive(turkey2) at 2.
s(2) :: xmasDay at 1.	s(2) :: chirp(bird) at 2.
s(2) :: gun(pjGun) at 1.	s(2) :: nearby(bird) at 2.
s(2) :: longWalk(pj) at 1.	s(2) :: aim(pjGun,turkey1) at 2.
s(2) :: animal(turkey1) at 2.	s(2) :: pulltrigger(pjGun) at 2.

The questions are answered after reading, respectively, the first and second story scenes:

session(s(1),[q(01)],all). session(s(2),[q(06)],all).

with their corresponding multiple-choice questions being:

q(01) ??
lives(pj,city) at 0; lives(pj,hotel) at 0; lives(pj,farm) at 0; lives(pj,village) at 0.
q(06) ??
motive(in(pj,forest),practiceShooting) at 3;
motive(in(pj,forest),huntFor(food)) at 3;
(motive(in(pj,forest),catch(turkey1)) at 3, motive(in(pj,forest),catch(turkey2)) at 3);
motive(in(pj,forest),hearBirdsChirp) at 3.

To answer question q(01), the system uses the following background knowledge:

p(11) :: has(home(pj),barn) implies lives(pj,countrySide).
p(12) :: true implies -lives(pj,hotel).
p(13) :: true implies lives(pj,city).
p(14) :: has(home(pj),barn) implies -lives(pj,city).
p(15) :: clean(pj,barn) implies at(pj,barn).
p(16) :: at(pj,home), at(pj,barn) implies has(home(pj),barn).
p(17) :: xmasEve, night implies at(pj,home).
p(18) :: working(pj) implies -at(pj,home).
p(111) :: lives(pj,countrySide) implies lives(pj,village).
p(112) :: lives(pj,countrySide) implies lives(pj,farm).
p(113) :: lives(pj,village) implies -lives(pj,farm).
p(114) :: lives(pj,farm) implies -lives(pj,village).
p(14) >> p(13). p(18) >> p(17).

By the story information, p(17) implies at(pj,home), without being qualified by p(18), since nothing is said in the story about Papa Joe working. Also by the story information, p(15) implies at(pj,barn). Combining the inferences from above, p(16) implies has(home(pj),barn), and p(11) implies lives(pj,countrySide). p(12) immediately dismisses the case of living in a hotel, whereas p(14) qualifies p(13) and dismisses the case of living in the city. Yet, the background knowledge cannot unambiguously derive one of the remaining two answers, with p(111), p(112), p(113), p(114) giving arguments for either. This is in line with the variability in the human answers to the first question.

To answer question q(06), the system uses the following background knowledge:

p(21) :: want(pj,foodFor(dinner)) implies motive(in(pj,forest),huntFor(food)).
p(22) :: hunter(pj) implies motive(in(pj,forest),huntFor(food)).
p(23) :: firedat(pjGun,X), animal(X) implies -motive(in(pj,forest),catch(X)).
p(24) :: firedat(pjGun,X), animal(X) implies -motive(in(pj,forest),hearBirdsChirp).
p(25) :: xmasDay implies want(pj,foodFor(dinner)).
p(26) :: longWalk(pj) implies -motive(in(pj,forest),practiceShooting).
p(27) :: xmasDay implies -motive(in(pj,forest),practiceShooting).

By the story and background knowledge parts not shown above, we can derive that Papa Joe is a hunter, and has fired at a turkey. From the first inference, p(22) already implies that the motivation is to hunt for food; the same inference can be derived by p(25) and p(21). At the same time, p(23) and p(24) dismiss the possibility of the motivation being to catch the two turkeys or to hear birds chirp, whereas story information along with p(26) or p(27) dismiss also the possibility of the motivation being to practice shooting.

An interesting example of variability occurred in the answers for the group of questions q(07), q(08), q(10), q(11), all asking about the status of the turkeys at various stages in the story. The majority of the readers followed a comprehension model which was revised to alternate between the first turkey being dead and alive. However, a minority of the readers consistently answered that both turkeys were alive. These readers had qualified the causal unit-argument supporting the conclusion that the first turkey was dead after being fired at, perhaps based on an expectation that the desire of Papa Joe for turkey would be met with complications. We believe that such expectations can be generated from standard story knowledge in the same way as we draw other elaborative inferences.

4. Related and Future Work

Automated story understanding has been an ongoing endeavor of AI for more than forty years [21,23]. Logic-related approaches largely proceed under the assumption that standard logical reasoning techniques can subsequently be applied; e.g., satisfiability [22] or planning [24]. To our knowledge very little work exists that relates story comprehension with computational argumentation, an exception being the work of Bex et al. [2,3], which combines narratives and argumentation in the context of legal reasoning. Our approach to base the development of an argumentation framework and systems for SC strongly on knowhow from psychology is novel. Argumentation for reasoning about actions and change, on which part of our formal framework builds, was studied in [17,28].

To complete a fully automated approach to SC we continue drawing lessons from psychology to address further the aspects of cognitive economy and coherence, by applying “computational heuristics” on top of our existing framework. We expect that psychology will guide us in modularly introducing operators such as *selection*, *dropping*, and *generalization* in order to implement a high-level of coherence in the computed comprehension models. We will also need to exploit more systematically knowledge on the general structure, content, and function of the story genre, as well as knowledge on reader expectations about characters and story plots [29]. This could be accommodated naturally within a preference-based argumentation framework by conditioning the argument priorities on readers’ expectations, thus dynamically changing as the story unfolds.

We are also investigating the systematic extraction / acquisition of world knowledge unit-arguments using lexical databases [1,19], knowledge archives [11], crowdsourcing

techniques [26], or machine learning on raw text (e.g., found on the Web) [14,15,16,18]. We envisage that the strong inter-disciplinary nature of our work can provide a concrete and important test-bed for evaluating the development of computational argumentation, while at the same time offering valuable feedback for psychology.

References

- [1] C. F. Baker, C. J. Fillmore, and J. B. Lowe. The Berkeley FrameNet Project. In *ACL*, 1998.
- [2] F. J. Bex, P. J. van Koppen, H. Prakken, and B. Verheij. A Hybrid Formal Theory of Arguments, Stories and Criminal Evidence. *AI and Law*, 18(2):123–152, 2010.
- [3] F. J. Bex and B. Verheij. Story Schemes for Argumentation about the Facts of a Crime. In *CMN*, 2010.
- [4] W. F. Brewer and E. H. Lichtenstein. Stories are to Entertain: A Structural-Affect Theory of Stories. *Journal of Pragmatics*, 6:473–486, 1982.
- [5] P. M. Dung. On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games. *AIJ*, 77(2):321–358, 1995.
- [6] A. Kakas and P. Mancarella. On the Semantics of Abstract Argumentation. *Logic Computation*, 23:991–1015, 2013.
- [7] A. Kakas, L. Michael, and R. Miller. Modular-E and the Role of Elaboration Tolerance in Solving the Qualification Problem. *AIJ*, 175(1):49–78, 2011.
- [8] A. Kakas, F. Toni, and P. Mancarella. Argumentation Logic. In *COMMA*, 2014.
- [9] W. Kintsch. The Role of Knowledge in Discourse Comprehension: A Construction-Integration Model. *Psychological Review*, 95:163–182, 1988.
- [10] W. Kintsch. *Comprehension: A Paradigm of Cognition*. NY: Cambridge University Press, 1998.
- [11] D. B. Lenat. CYC: A Large-Scale Investment in Knowledge Infrastructure. *CACM*, 38(11):32–38, 1995.
- [12] D. S. McNamara and J. Magliano. Toward a Comprehensive Model of Comprehension. *The Psychology of Learning and Motivation*, 51:297–384, 2009.
- [13] H. Mercier and D. Sperber. Why Do Humans Reason? Arguments for an Argumentative Theory. *Behavioral and Brain Sciences*, 34(2):57–74, 2011.
- [14] L. Michael. Reading Between the Lines. In *IJCAI*, 2009.
- [15] L. Michael. Causal Learnability. In *IJCAI*, 2011.
- [16] L. Michael. Machines with Websense. In *Commonsense*, 2013.
- [17] L. Michael and A. Kakas. Knowledge Qualification through Argumentation. In *LPNMR*, 2009.
- [18] L. Michael and L. G. Valiant. A First Experimental Demonstration of Massive Knowledge Infusion. In *KR*, 2008.
- [19] G. A. Miller. WordNet: A Lexical Database for English. *CACM*, 38(11):39–41, 1995.
- [20] S. Modgil and H. Prakken. A General Account of Argumentation with Preferences. *AIJ*, 195:361–397, 2012.
- [21] E. T. Mueller. Story Understanding. In L. Nadel, editor, *Encyclopedia of Cognitive Science*, volume 4, pages 238–246. London: Macmillan Reference, 2002.
- [22] E. T. Mueller. Story Understanding through Multi-Representation Model Construction. In *HLT-NAACL 2003 Workshop on Text Meaning*, pages 46–53, 2003.
- [23] E. T. Mueller. Story Understanding Resources. <http://xenia.media.mit.edu/~mueller/storyund/storyres.html>, 2013. Accessed February 28, 2013.
- [24] J. Niehaus and R. M. Young. A Computational Model of Inferencing in Narrative. In *INT*, 2009.
- [25] D. N. Rapp and P. Van den Broek. Dynamic Text Comprehension: An Integrative View of Reading. *Current Directions in Psychological Science*, 14:297–384, 2005.
- [26] C. Rodosthenous and L. Michael. Gathering Background Knowledge for Story Understanding through Crowdsourcing. In *CMN*, 2014.
- [27] M. Thielscher. The Qualification Problem: A Solution to the Problem of Anomalous Models. *AIJ*, 131(1–2):1–37, 2001.
- [28] Q. B. Vo and N. Y. Foo. Reasoning about Action: An Argumentation-Theoretic Approach. *JAIR*, 24:465–518, 2005.
- [29] R. A. Zwaan. Effect of Genre Expectations on Text Comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20:920–933, 1994.