Advances in Cognitive Systems 5 (2017) 1-6 Submitted 3/2017; published X/20XX

Character-building stories

Dylan Holmes DXH@MIT.EDU
Patrick Winston PHW@MIT.EDU
Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA, 02139 USA

Abstract

We argue that story understanding mechanisms provide a foundation for modeling aspects of our ability to reason hypothetically. We first note that story understanding mechanisms enable us to answer what-if questions about what would happen if an event did or did not occur, and we note that story understanding enables us to answer what-if questions about how a story would be interpreted from a different cultural perspective. We then advance a theory of how humans use hypothetical reasoning to think about personality traits. Our theory and implementation describe how humans use past behavior and untapped alternatives to build a model of characters' motives and constraints. We focus on how generalizations of existing story understanding methods and concepts enable us to model this competence efficiently. In a sample story, our theory and implementation perform a complex reasoning process to decide what a character will do next based on whether the character is more like a Conformist, Thief, Opportunist, or Robin Hood archetype.

1. Vision

Varieties of hypothetical reasoning pervade intelligent behavior [Sloman, 2015]. From low-level visual processing where we perceive which surfaces are graspable [Gibson, 1979], to high-level legal argument where we reason about self-defense in terms of harm that might have happened [Rissland, 1989], to engineering tasks where we anticipate potential failure modes or write programs that introspect on their reasons for acting [Forbus and Hinrichs, 2006], much of what we do depends in some form on our ability to think in terms of possibility, impossibility, and constraint.

As a special case, hypothetical reasoning enriches our comprehension of stories. Consider how the capacity for suspense, surprise, and poignancy depend on the capacity to imagine what is possible ("What would happen if Romeo had learned Juliet's death was a ruse?").

We propose that a converse of this idea is also true: the skills that enable us to understand stories serve as a fruitful foundation for the capacity to imagine and manipulate alternative circumstances—that is, to reason hypothetically in a particular sense. We have found that from an engineering perspective, story understanding aids in hypothetical reasoning, and believe that our work sheds light on the science of human performance as well.

In previous work, we have shown how our existing framework of story-understanding mechanisms, the Genesis story-understanding system, can enable us to argue hypothetically about self-defense ("What would happen if Alex didn't brandish a knife?") or reason from a hypothetical

point of view in a story about the 2007 Russia-Estonia cyberwar ("What would happen if I were from Estonia?") [Holmes and Winston, 2016].

In this paper, we apply our story-enabled hypothetical reasoning approach to the problem of modeling personality, goal-directed behavior, and moral constraint. We propose a theory of how humans reason in this way, and show how we can efficiently model this competence using story-understanding capabilities. The program we have developed reads short stories in simplified English, building models of each character. When prompted with a question from the user like "What would happen if Amy wanted the robot?", the program uses those character models to predict a character's future response to a novel scenario.

Our aims in this paper are twofold. First, we aim to highlight the variety of hypothetical reasoning processes underlying our cognitive abilities, processes which range from perception of visual affordances to moral judgments to block-building plans. From a scientific standpoint, we believe that hypothetical reasoning forms a useful focal point for studying cognitive processes. From an engineering standpoint, we believe that thinking in terms of possible alternatives can drive us to develop broader and more robust cognitive systems.

Second, we aim to exhibit hypothetical reasoning ability as an instance of our *strong story hypothesis* [Winston, 2011]: We believe that the mechanisms that enable us to understand stories distinguish human intelligence from the intelligence of other species, and as a corollary, propose that these story-understanding mechanisms enable our powerful human ability to reason about possibility, impossibility, and constraint.

2. We analyze how humans judge goals and personality

Our aim is to model our human facility with personality and goal-directed behavior. The central focal point is the following story and associated question:

> Amy is at the playground. Jeff is playing with the ball. Amy asks Jeff for the ball, so Jeff gives the ball to Amy. Amy plays with the ball. Teresa steals the ball from Amy and plays with the ball. Then, Amy goes to the cafeteria. Kate is Amy's friend. Kate doesn't have food. Amy steals food from the cafeteria and gives it to Kate. Then, Amy walks home. Amy passes a toy store. The toy store has a robot.
>
> *What would happen if Amy wants the robot?*

In order to enable a computer to answer such questions, we must analyze how we ourselves answer them. Here, we present our analysis: in the first stage, it seems we make a reflexive judgment based on precedent. (In an informal survey, a majority suggested that Amy might steal the robot given that she had stolen before.) Then, if sufficiently motivated, we engage a more deliberative problem-solving process. The apparent purpose of this deliberative process is to uncover additional evidence pertaining to the question ("consider Amy's motive in stealing food and the fact that she does not steal the ball from Jeff"). These evidence-gathering procedures often weigh counterfactual alternatives and highlight implicit moral constraints, which is to say they are often hypothetical. All of this evidence seems to be in service of forming a mental model of people in terms of their

motives, their methods, and their moral constraints. Such a personality model is then used to predict behavior and answer the original query.

Delving deeper into personality models, we ask how we arrive at the specific models we do. One common judgment about Amy, for example, is that she is what we might call a *Robin Hood* type character: though she steals some of the time, it is never for personal gain. Adopting this judgment for the sake of argument, we can contrast this judgement with several other potential personality ascriptions. Why do we not consider Amy to be an inveterate thief, stealing whatever she wants—after all, we have positive evidence that Amy *does* steal food—and what would constitute opposing evidence? Why should we bother to conclude that Amy is operating under a moral constraint (avoiding theft for personal gain on principle), rather than simply acting opportunistically, stealing whenever the mood strikes? Finally, how do we recognize that particular incidents (such as Amy asking Jeff for the ball, or Amy stealing food from the cafeteria) are *relevant* to the question in the first place? After all, the question asks what would happen if Amy *wants* the robot, yet at a superficial textual level, neither of the words 'want' nor 'have' appear anywhere in the story.

We propose the following theory:

1. **How do we recognize relevant incidents?** We link the question and the story through knowledge of means and ends. In this particular story, we look for all actions that have *acquisition* as an implicit goal. This allows us to link Amy's *wanting* the robot, *asking* for Jeff's ball, and *stealing* food from the cafeteria. Characters' previous actions reveal their goals, their methods, and their moral constraints (or lack thereof).

2. **What do we consider when we deliberate?** We dig up additional evidence, often weighing counterfactual alternatives and highlighting implicit moral constraints. We use this (often hypothetical) evidence to form a mental model of people's motives, methods, and moral constraints—an aspect of *personality*. (Note: in this paper, we are considering just the aspects of personality which pertain to motives, methods, and moral constraints..) We use personality models to predict behavior and answer the original query.

3. **Which personalities fit best?** We assign personalities by aligning characters with known character archetypes that fit best. To assign personalities, we must have methods for evaluating and comparing different personality types.

   In this story, to see Amy as a Robin Hood character (who steals but never for personal gain), we must rule out other apparently plausible accounts: for example, that Amy is simply a thief (as precedent suggests), or that Amy is an opporunist (operating without any moral constraint).

   - **Why not just think that Amy is a thief?** We note that Amy asked for Jeff's ball, rather than stealing it. When a character achieves the same kind of goal through different means in different situations, we may resist a one-sided characterization.
   - **Why believe that Amy operates under constraint?** Amy steals food to benefit a friend, but does not steal a ball to benefit herself. When a character could have chosen a constraint-violating strategy but did not, we may infer that the character heeds the constraint.

Hence, a small number of principled heuristics like these can guide our sense of fit. These heuristics crucially consider hypothetical alternatives to rule out (or promote) certain personality types.

4. **How do we predict behavior using personalities?** Once we know a character's goals, methods, and constraints, we can simulate their possible moves and eliminate the forbidden. When we decide that Amy is a Robin Hood character, we conclude that Amy would refuse to steal the robot, and would be more likely to ask for it instead.

Using the story-understanding framework provided by Genesis, we have built a program, which we call PERSONATE, that implements this theory and models the behavior of a reader weighing evidence and evaluating personalities in order to predict a character's actions (Figure 1).

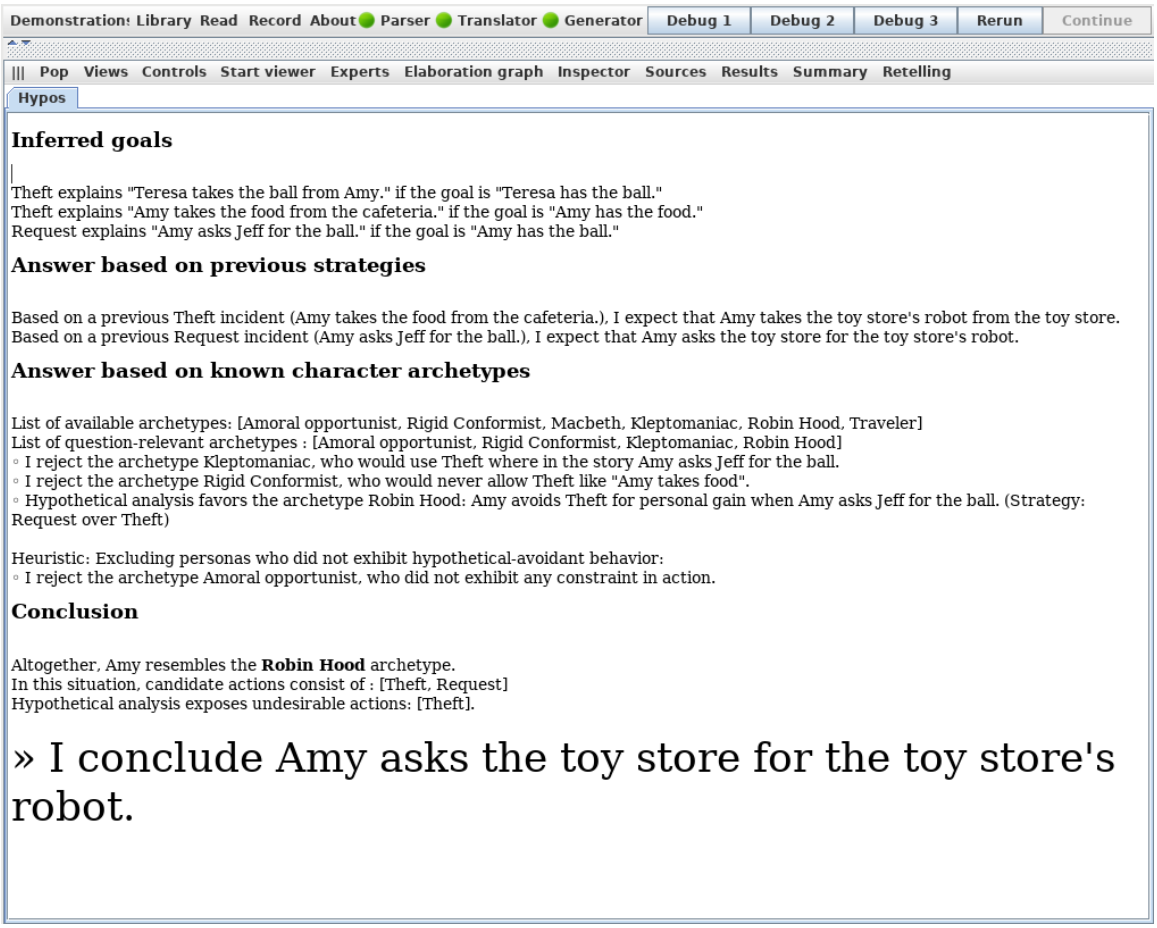

Figure 1. A complete trace of PERSONATE reading Amy's story and predicting future behavior. Personality-based analysis consists of several stages of problem solving, many of which employ hypothetical reasoning.

In the next section, we describe the implementation of PERSONATE and how we were able to develop this whole new hypothetical reasoning capability efficiently by building upon a foundation of existing story understanding capabilities.

3. PERSONATE predicts behavior using goals and personalities

In this section, we describe our implementation of the theory outlined in the previous section. The subsections mirror the principles enumerated above. In particular:

§3.1 **How do we recognize relevant incidents?** We look for goal-directed behaviors in the story. PERSONATE uses a database of means-ends rules to infer character goals from character actions. PERSONATE limits search by considering only goals that match the user's query.

§3.2 **What do we consider when we deliberate?** We take our knowledge of the methods each character has previously used and speculate about how the character has selected one method over another—their constraints, preferences, decision procedures and so on. PERSONATE has a library of archetypal *personas* which are bundles of means-ends rules along with constraints in the form of *forbidden concept patterns*. For each character, PERSONATE produces a list of loosely-matching candidate personas. The list is refined in subsequent steps.

§3.3 **Which personalities fit best?** We use heuristics to evaluate and compare candidate personas. In particular, PERSONATE uses just four heuristics: check forbidden concepts, reject oversimplified personas, reward actively avoided constraints, and prefer parsimony. These four heuristics, several of which involve hypothetical reasoning about alternatives, capture important aspects of the way we intuitively decide which personality descriptors fit.

§3.4 **How do we predict behavior using personalities?** Having eliminated unlikely candidate personas, PERSONATE uses the remainder to predict behavior. In the case where one persona remains, for example, PERSONATE predicts that the character may use any of the available strategies that do not violate the persona's constraints. PERSONATE uses hypothetical reasoning to anticipate constraint-violating side effects of each action.

3.1 We recognize relevant incidents using Means-and-ends knowledge

We must first explain how we humans decide what information in the story is relevant to the question "What would happen if Amy wants the robot?" In particular, we must explain our intuition that the sentences "Amy steals food from the cafeteria" and "Amy asks Jeff for the ball" are both relevant, while a syntactically similar sentence such as "Amy rides home on a bike" would not be.

Our solution is that we recognize relevant incidents using knowledge of means and ends. To represent such knowledge in PERSONATE, we developed *means-ends rules*. Means-ends rules are a variant of Genesis's heuristic inference rules, which Genesis uses to supply missing common-sense facts, inferences, and connections between events in a story [Winston, 2014].

Means-ends rules likewise supply common-sense information. A means-ends rule comprises a goal, a list of prerequisites, and a means (action). For example, this simple scenario depends on two such rules:

```
Strategy: "Theft"
Goal: xx has zz
Prerequisites: yy has zz
Means: xx takes zz from yy

Strategy: "Request"
Goal: xx has zz
Prerequisites: yy has zz
Means: xx asks yy for zz
```

Semantically, these means-ends rules encode commonsense knowledge about what characters may do (Means) depending on their desires (Goal) and present situation (Prerequisites).

Thus, by matching the question "What would happen if Amy wants the robot?" against the means-ends database, PERSONATE can populate a list of logically possible answers: using Theft, Amy might steal the robot; using Request, Amy might ask for it. (Detail: An idiomatic transformation finesses "Amy wants the robot" into "Goal: Amy has the robot".)

PERSONATE can also identify means and ends in the story itself. This enables PERSONATE to give not only logically reasonable answers, but answers supported by precedent in the story. Thus, by matching against Means and binding to Goals, PERSONATE can guess what characters want and how they get it: presumably, Amy asks for the ball because Amy wants [to have] the ball; Teresa steals the ball because Teresa wants the ball; and Amy steals food from the cafeteria because Amy wants the food[1]. In this way, PERSONATE obtains a guess about what characters want and observations about what methods they use to get it. Thus, PERSONATE can hazard a guess about what Amy will do if Amy wants the robot, citing precedent (and implicit goals):

> Amy may steal the robot because Amy stole food from the cafeteria.
> Amy may ask for the robot because Amy asked Jeff for the ball.

3.2 Deliberation uncovers personality: means, methods, and moral constraints

PERSONATE's reflexive predictions, like our own, are often rather crude: though means-ends analysis reveals a character's library of previously-used actions, PERSONATE has no principle to decide *which* method the character will use.

To remedy this, in the deliberative stage, PERSONATE imitates our human ability to draw on (often hypothetical) evidence to build a better model of character. This improved model supplements the character's library of available actions with the *constraints*—preferences, allegiances, morals—that govern which methods the character uses in different situations. For our purposes, we explore constraints of a particular form: **forbidden concept patterns**.

In the Genesis system, a *concept pattern* is a constellation of events in a story. Many, but not all, concept patterns involve *leads-to* relationships; that is, relationships that emerge from an unbroken chain of events and inferences in a story. For example, the concept pattern Revenge occurs whenever

1. This last inference is, presumably, untrue, as Amy is stealing food on Kate's behalf. However, PERSONATE's reflexive assignment of goals need not be correct.

one act of harm is connected to a reciprocal act of harm through any number of intervening story elements (Figure 2).

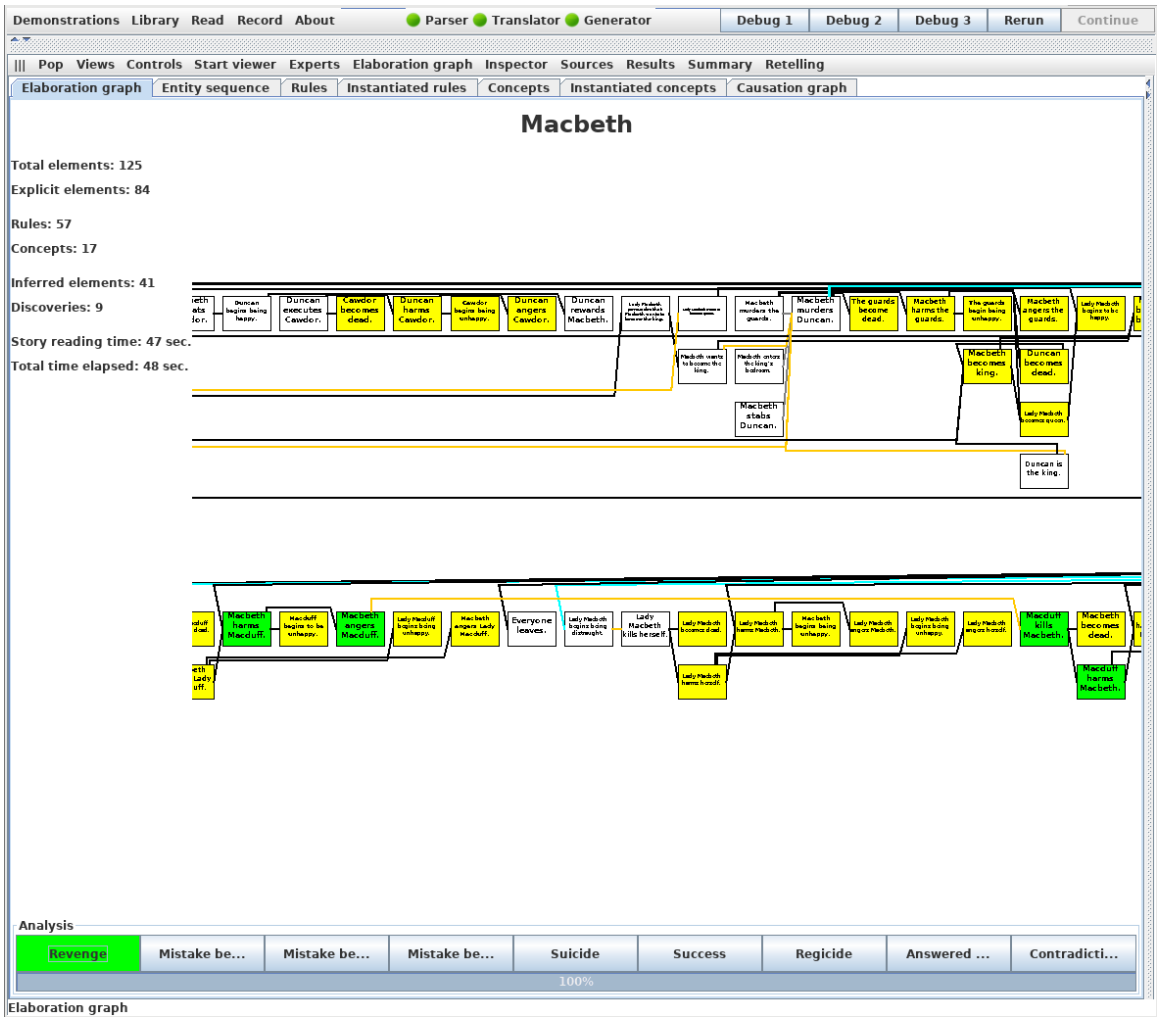

Figure 2. The *elaboration graph* shown here depicts the events in a simplified version of *Macbeth*, including deduced facts and conjectured causal connections. Concept patterns such as Revenge (highlighted in green) emerge from the long-distance chains of such causes or inferences in the narrative.

For PERSONATE, we borrow the existing concept-pattern apparatus as a way of identifying constraints. For example, we define a concept pattern "Theft for personal gain" which is rendered as xx steals zz from yy eventually leads to xx enjoying zz; we also define the simple pattern "Law-breaking", which occurs whenever xx steals. By associating such concept patterns with particular characters, we can define a personality type that refuses to steal for personal gain, or one that refuses to steal at all, respectively.

A *persona* is our implementation of such a personality model. Personas are a simple extension of Genesis's existing mental models, collections of rules and concept patterns that define a particular reader's commonsense background [Winston, 2014]: each persona comprises a library of available methods (in the form of means-ends rules) and behavioral constraints (in the form of forbidden concept patterns.) To demonstrate our personality-choosing heuristics, we supply PERSONATE with four candidate personas. (In fact, we have defined additional personas, but because these are not related to acquisition, PERSONATE disregards them when answering the question "What would happen if Amy wants the robot?". For example, the Macbeth persona embodies the strategy of regicide to become king. The Pathfinder persona embodies the strategy "if your goal is to be in location xx, then go to xx." Furthermore, although the list of personas is presently hardcoded, we can envision a variation where character archetypes are distilled from previously read stories.)

```
Persona: "Conformist"
Means-ends: Theft, Request
Forbidden concepts: Lawbreaking.

Persona: "Thief"
Means-ends: Theft
Forbidden concepts: None.

Persona: "Opportunist"
Means-ends: Theft, Request
Forbidden concepts: None.

Persona: "Robin Hood"
Means-ends: Theft, Request
Forbidden concepts: Theft for personal gain.
```

As a point of clarification, we note that a persona's means-ends strategies include all strategies that the persona *knows*, not necessarily the ones the persona will use. For example, the Conformist knows the Theft strategy but will never use it. The Thief represents a kind of child-like person whose only known method for acquiring something is stealing it.

In the next section, we show how PERSONATE makes a reasoned human-like argument, using hypothetical reasoning heuristics to reject all but one persona from this list. Such reasoning ultimately leads PERSONATE to conclude that Amy best resembles the Robin Hood persona, laying the groundwork for a behavioral prediction and an answer to the query "What would happen if Amy wants the robot?".

3.3 A small number of principled heuristics determine personality fit

Now we must explain how PERSONATE, and likewise human beings, decide which personality types fit best. In this story, for example, we intuitively feel that Amy is more like a Robin Hood character (stealing only to benefit others), rather than an inveterate thief (as precedent suggests) or

an opportunist who acts without moral constraint. In this section, we describe how PERSONATE captures intuitive assessments like these using a small number of principled heuristics for promoting or eliminating candidate personas. These heuristics often centrally involve hypothetical reasoning about characters' available alternatives and avoided outcomes.

To start, PERSONATE populates an initial list of candidate personas. This list of candidates is extremely permissive: it includes any personality having a means-end rule that is employed in the story. For example, in the previous stage, PERSONATE detected the following means-ends rules in the story: Amy gets the ball from Jeff by asking for it (Request); Teresa gets the ball from Amy by stealing it (Theft); Amy gets the food from the cafeteria by stealing it (Theft). Hence, in this stage, PERSONATE will initially consider any personas that employ the Theft or Request strategies. However, PERSONATE avoids doing too much unnecessary work by ascribing a personality only to the person mentioned in the query—a kind of question-directed search. For our story, this search process yields the four candidate personas noted above: Conformist, Thief, Opportunist, Robin Hood. (Each of these personas employs a strategy used by Amy during the story.)

Next, we heuristically shorten this list by eliminating unlikely candidates. This process resembles a kind of near-miss learning of personality type, using counter-examples in the story to hone an emerging model. We use Amy's story as a concrete example to demonstrate our personality-evaluating heuristics in practice.

Of the four candidate personas in our story, the most straightforward to eliminate is the Conformist (who never breaks the law): our human intuition is that Conformist is a bad fit because Amy steals food from the cafeteria. PERSONATE's corresponding heuristic is to eliminate all candidate personas whose forbidden concept pattern (here, Lawbreaking) appears explicitly in the story.

Given that Amy does steal food from the cafeteria, should we conclude that Amy is simply a thief who steals whatever she wants? Our intuition is to resist such a one-sided characterization. As justification, we might cite the fact that Amy gets the ball from Jeff by asking for it—she does not steal the ball in that case, although she could have. PERSONATE's corresponding heuristic is to consider all methods that a character employs in service of each particular goal. Then, if a character knows more methods for achieving a goal than the candidate persona does, we reject the persona as being too simplistic. (In this case, PERSONATE rejects the Thief persona, which cannot account for how Amy gets the ball from Jeff by asking for it.) Note that this deliberative process imitates how humans, having made a reflexive judgement based on means-ends precedent, later reflect on the character's other actions and adopt a more nuanced characterization.

Now we consider why we might be justified in explaining Amy's behavior as operating under a Robin Hood constraint (stealing only to help others), rather than simply doing whatever she wants. To be sure, both remaining candidate personas—Robin Hood and the Opportunist—account for Amy's previous behavior equally well. That is to say, both contain all of the means-ends strategies Amy employed during the story. Moreover, the Opportunist is arguably a simpler model, as it includes no constraints. Why conclude that Amy is deliberately avoiding theft for personal gain?

Our intuition is that Amy steals *only* to benefit others: for one thing, Amy steals food from the cafeteria for Kate, not herself. More interestingly, Amy gets the ball from Jeff by asking rather than stealing it: when she wants something for *herself*, evidently Amy does not steal it. This kind of argument requires hypothetical reasoning, thinking about the actions Amy *could* have taken

and what their consequences would have been. PERSONATE's corresponding heuristic is that if a character's unused alternatives trigger a persona's forbidden concept pattern, PERSONATE prefers that persona. To this end, PERSONATE considers all the means-ends strategies the character uses in the story. For each strategy, PERSONATE considers other known methods of achieving that same goal. For each alternative method, PERSONATE checks whether that alternative method would have activated a forbidden concept pattern. If so, PERSONATE concludes that the character may have avoided the forbidden concept *on principle*. When we find that a persona has complicated constraints, and yet we note character's choices adhere to those constraints, we consider the persona to be a more falsifiable, more predictive model of behavior. We prefer personas with more "actively avoided" constraints in the story. In this case, Amy avoids the "Theft for personal gain" concept pattern when asking Jeff for the ball rather than stealing it, so PERSONATE prefers the Robin Hood characterization over the Opportunist.

This concludes our analysis of Amy: using these general heuristics, PERSONATE decides that Amy is most like a Robin Hood character. As a final addendum, we demonstrate how these heuristics would interact in other stories and other questions. For example, consider the same story with the question "What would happen if *Teresa* wants the robot?" The dedicated reader will confirm that PERSONATE will initially consider all four personas: Conformist, Thief, Opportunist, Robin Hood. Because Teresa steals the ball for herself, Teresa exhibits the concept pattern "Lawbreaking" and "Theft for personal gain"; thus, Teresa cannot be a Conformist or a Robin Hood character (according to the first heuristic.) Instead, Teresa must either be a Thief or an Opportunist — none of the remaining heuristics can help us choose between them. At this point, we may reasonably decide that we lack enough evidence to choose one persona over the other. If pressed to pick only one persona, however, we might consider using a parsimony heuristic: prefer models with fewer components; that is, fewer means-ends rules and fewer concept patterns. This heuristic, potentially undesirably, would prefer characterizing Teresa as a thief rather than an opportunist.

To summarize this section, PERSONATE implements our human judgements using the following heuristics:

1. **Check forbidden concepts**. If a character participates in a persona's forbidden concept pattern, reject that persona.
2. **Reject oversimplified personas**. If a character knows more means to the same end than a persona, reject that persona.
3. **Reward actively avoided constraints**. If a character's unused alternatives trigger a persona's forbidden concept pattern, prefer that persona. The more avoided concepts, the better.
4. **Prefer parsimony as a constraint of last resort**. When pressed, you may prefer personas with fewer means-ends rules and fewer constraints.

The second and third heuristics are explicitly hypothetical. They consider alternative methods for achieving the same end and evaluate the consequences of those alternatives.

3.4 Models of personality circumscribe behavior

We have now described how PERSONATE identifies goal-seeking behavior in the story and uses a suite of heuristics and hypothetical reasoning capabilities to integrate those behaviors into a model of personality. Intuitively, once we have a model of the character's personality, we can use that model to predict behavior and answer questions such as "What would happen if Amy wants the robot?". The character's available methods constitute possible actions, and the character's constraints help us determine which of those actions the character will choose in a novel situation.

PERSONATE's corresponding behavior starts with the finalized list of candidate personas produced in the previous section. If there is only one candidate remaining, the next step is straightforward: consider every method (means-ends rule) the persona has for achieving the goal in question. If any of those methods activate the persona's forbidden concept patterns, eliminate them. Report that the character may use any of the remaining means. For example, in our story, Amy fits the Robin Hood persona best. The Robin Hood persona has two methods for acquiring the robot: stealing it, or asking for it. Stealing it would constitute "Theft for personal gain", hence PERSONATE concludes that Amy will not steal the robot. PERSONATE reports that, upon reflection, Amy will ask for the robot instead. This represents the culmination of PERSONATE's ability to predict behavior from personality.

If there is more than one candidate persona remaining, it is less clear what PERSONATE ought to do. For our purposes, PERSONATE uses a straightforward, if cumbersome, generalization of the one-persona strategy. PERSONATE considers all possible methods available to the remaining personas and eliminates those that are constraint-violating for every remaining persona. PERSONATE reports that all remaining methods are possible. In future work, we can imagine PERSONATE using more sophisticated principles to choose between strategies.

4. Discussion and Contributions

In this paper, we have developed a theory and a program (PERSONATE) which models human reasoning about goal-directed behavior and constraint in terms of hypothetical alternatives. We introduced a technical sense of the term "persona" or "personality" to refer to these clusters of means-ends rules and associated constraints; PERSONATE answers questions about predicted behavior by assigning such personas to characters in the story. The everyday term 'personality' (along with nearby terms such as 'archetype', 'character', and 'moral values') presents some difficulties, as it is overloaded with potentially many different senses—it is a *suitcase word* [Minsky, 2006, Chapter 4] or a *cluster concept* [Sloman, 2011].

In this paper, we use the term personality specifically to refer to enduring (typically moral) constraints coupled with goal-seeking behavior, and our aim has been to develop a model of how hypothetical reasoning can help elucidate characters' goals and constraints by, for example, highlighting the actions they could have otherwise taken.

Hence the focus of our work in this paper complements, but contrasts with, other approaches and other senses of the term "personality". Unlike Digman's (1990) five-factor model of personality, for example, we do not attempt to model behavior at its highest level of abstraction as a few named dimensions of variability. Instead, we attempt to partially model the cognitive decision process, and

our models are so fine grained as to be specialized to a particular purpose (such as acquisition). We envision that fuller descriptions of characters would involve *collections* of such goal-specific personas, which characters would variously exhibit based on task and changing mood. (We anticipate that regularities in the personas employed by a particular character could comprise a useful, more abstract, signature of personality.)

Unlike Rizzo, Veloso, Miceli, and Cesta (1997), we do not attempt to distinguish character types in terms of the kind of goals they are likely to have (such as keeping interpersonal commitments or experiencing entertainment). Though our work is similarly goal-based, we place emphasis on how personality is defined by constraints revealed by choices between alternative means of achieving goals: in our case, the goals themselves are not especially characteristic of personality, whereas the choice of means is. Finally, with this program, we consider only simplified goals that might be achieved in one step: we use planning and search not to determine the activities of agents, but to decide what evidence to investigate in the story and what hypothetical variations would best expose constraint and clarify personality.

As we saw in Figure 1, PERSONATE infers character goals from character actions, makes predictions based on precedent, and refines predictions using partial models of personality (personas). The procedure for evaluating and assigning personas is fundamentally hypothetical. We introduced means-ends rules (which might be considered analogous to If-Do-Then rules [Minsky, 2006, Chapter 4] or reminiscent of STRIPS operators, except insofar as our means-ends rules are expressed as English sentences with variables and are not involved in chains of actions.) Within our theory, we capture important aspects of our reasoning process using a few ideas: Means-ends rules expose motivation, question-directed problem-solving reduces search workload, characters align with personas that capture constraint, key incidents in the story offer near-miss examples for learning personas, and four heuristics help evaluate and compare personas.

Though we focused our analysis on a single distilled scenario which shows off the key capabilities of our system, the ideas we put forth are flexible and extensible: the library of Means-ends rules can be enlarged to include different types of goal-directed behavior, and even a larger library can be indexed efficiently by goal; personas can be applied to other goal-directed behavior, and in future work could be distilled from stories; our four heuristics capture important and general features of how we evaluate personalities, and so can be applied to other stories or extended as we learn more about how people reason. Of course, we make no assumptions that scaling up to hundreds of unseen stories and dozens of personas will be an easy engineering task. Indeed, the *learning* problem—the problem of discovering new problem-solving methods and constraints, which we do not address here—can involve aggregating a variety of contingent, commonsense facts that people deploy when solving their problems. Nevertheless, we believe that the conceptual framework of goals, constraints, and personas which we develop in this paper, in harness with a story-understanding substrate as provided by Genesis, provides a powerful language for expressing these kinds of goal-directed behaviors.

We have proposed that the mechanisms that enable us to understand stories form a powerful foundation for our ability to reason hypothetically, and that varieties of hypothetical reasoning underlie much of our intelligent behavior. At the heart of our program, two key ideas emerge:

1. We learn the most about a character from the actions they could have taken, but avoided.

2. The mechanisms that enable us to understand stories enable us to grasp such hypothetical possibilities.

References

John M Digman. Personality structure: Emergence of the five-factor model. *Annual review of psychology*, 41(1):417–440, 1990.

Kenneth D. Forbus and Thomas R. Hinrichs. Companion cognitive systems: a step toward human-level ai. 2006.

James J. Gibson. *The Ecological Approach to Visual Perception*. 1979.

Dylan Holmes and Patrick Winston. Story-enabled hypothetical reasoning. In *Proceedings of the Fourth Annual Conference on Advances in Cognitive Systems*, June 2016.

Marvin Lee Minsky. *The Emotion Machine*. 2006.

E. L. Rissland. Dimension-based analysis of hypotheticals from supreme court oral argument. In *Proceedings of the 2Nd International Conference on Artificial Intelligence and Law*, ICAIL '89, pages 111–120, New York, NY, USA, 1989. ACM. ISBN 0-89791-322-1. doi: 10.1145/74014.74030. URL http://doi.acm.org/10.1145/74014.74030.

Paola Rizzo, Manuela Veloso, Maria Miceli, and Amedeo Cesta. Personality-driven social behaviors in believable agents. In *Proceedings of the AAAI Fall Symposium on Socially Intelligent Agents*, pages 109–114, 1997.

Aaron Sloman. Family resemblance vs polymorphism, 2011. URL http://cs.bham.ac.uk/cogaff/misc/impossible.html. Accessed: 2017-05-03.

Aaron Sloman. Impossible objects, 2015. URL http://cs.bham.ac.uk/cogaff/misc/impossible.html. Accessed: 2017-02-23.

Patrick Henry Winston. The strong story hypothesis and the directed perception hypothesis. *AAAI*, 2011.

Patrick Henry Winston. The genesis story understanding and story telling system: A 21^{st} century step toward artificial intelligence. Memo 019, Center for Brains Minds and Machines, MIT, 2014.