Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT
Artificial Intelligence 153 (2004) 287–305
Artificial Intelligence
www.elsevier.com/locate/artint

The representation of planning strategies

Andrew S. Gordon

The Institute for Creative Technologies, The University of Southern California, 13274 Fiji Way, Marina del Rey, CA 90292, USA

Received 14 November 2001

Abstract

An analysis of strategies, recognizable abstract patterns of planned behavior, highlights the difference between the assumptions that people make about their own planning processes and the representational commitments made in current automated planning systems. This article describes a project to collect and represent strategies on a large scale to identify the representational components of our commonsense understanding of intentional action. Three hundred and seventy-two strategies were collected from ten different planning domains. Each was represented in a pre-formal manner designed to reveal the assumptions that these strategies make concerning the human planning process. The contents of these representations, consisting of nearly one thousand unique concepts, were then collected and organized into forty-eight groups that outline the representational requirements of strategic planning systems.

Keywords: Knowledge representation; Commonsense reasoning; Mental models; Knowledge-based planning; Planning strategies

1. Representation in models of planning

One of the hallmarks of Artificial Intelligence planning systems is the general absence of representational commitments. In many ways, this line of research has been successful in achieving its goal of building a *general problem solver* [10], where this has come to mean an algorithm that searches for a successful solution to a problem (a plan) given any well-specified search space. Over the many years that researchers have developed these algorithms, only a few representational constraints have been needed to advance the technology, notably the introduction a planning operator formalism [6], a notation for

E-mail address: gordon@ict.usc.edu (A.S. Gordon).

0004-3702/$ – see front matter
doi:10.1016/j.artint.2002.11.002

specifying quantified preconditions and context-dependent effects [12], and a means of describing plans as a hierarchy of tasks [13]. The strength of this research tradition is that it has enabled comparative evaluations of implemented planning systems, where each is given identical specifications of the search space [9].

The problematic downside to the representational agnosticism of artificial intelligence planning systems is the gulf that it creates between these technologies and the real humans that would like to use them for everyday problem solving and planning. Highlighted by continuing work on mixed-initiative planning (e.g., [5]), human-computer cooperation will require the development of systems with representational commitments that are more in accordance with people's commonsense understanding of goal-directed behavior. The difficulty in achieving this accord stems from the apparent breadth of our commonsense understanding of intentional action, which includes slippery concepts of commitment, preferences, opportunities, threats, ability, prevention, enablement, failures, constraints, decisions, postponement, interference, and uncertainty, among many others.

The challenge of describing human models of planning has been taken up by researchers in the field of commonsense reasoning, who have produced a body of work under the umbrella label of reasoning about rational action. This research is conducted within the paradigm of formal logic, where the aim is to author statements and axioms that can be employed in deductive inference systems (e.g., [11]). As an approach for investigating human commonsense reasoning, the key assumption is that formal theories describe our cognitive representations when the provable theorems of these theories match with our natural human inferences [3]. Leaving aside potential problems with this assumption, this research approach suffers most in that it scales up very poorly in practice. Consider the amount of research effort required to justify the inclusion of any single commonsense planning concept within a larger theory, perhaps best exemplified in the formal treatment of the concept of *intention* [1]. First, it was recognized that there were problems with existing formal theories of planning having to do with intention. Then, researchers authored a new modal logic to enable the expression of a set of axioms that included a handful of new concepts advocated among epistemological philosophers. Finally, it is shown how the inclusion of these new concepts within this new logic enables the proofs of certain theorems that more closely align with our commonsense inferences about planning than did the previous theories.

The significance and impact of this particular work on intention can hardly be understated in light of the enormous attention that it continues to receive. What is in question, however, is whether this work is an appropriate model for how the investigation of representation in commonsense reasoning should proceed. If we imagine that a complete commonsense model of human planning will be on the order of a few dozen concepts and axioms, then this approach is well on its way to success. If, instead, it is on the order of hundreds, thousands, or even millions of such formalizations, then the scope of the endeavor is probably too great for this scientific community to support. As we have had no means of estimating the scale of such a model, the inclination is to pursue alternative research methodologies. Is there a more direct approach to describing the representational components of human commonsense reasoning about planning?

This article explores an alternative methodology for the investigation of representation in commonsense models of human planning, one that capitalizes on the curious nature of *strategies*—in the common sense of the word. The notion of strategy takes on paradoxical roles in our intellectual lives. On one hand, strategies are the currency of everyday problem solving—we seem to have an effortless ability to employ and recognize a myriad of strategies in our occupational, recreation, and personal pursuits. On the other hand, individual strategies are uncomfortably abstract, seemingly impenetrable concepts that defy simple attempts at classification or definition. However, to a knowledge engineer, an individual strategy poses a particularly interesting puzzle, one where it is impossible to author a strategy definition without making certain representational commitments about the reasoning processes of the agent executing the strategy.

Consider an example of a strategy used by a budding concert pianist, as he later described it. During one of his early competitive performances his execution of a solo piano piece was derailed by an audience full of distractions, which included audible conversation, chair noises, and even the ringing of a mobile telephone. Determined to avoid this problem in the future, he would purposefully cause distractions to occur during his practice sessions. By setting alarm clocks for odd times, turning on kitchen oven timers, and asking some friends to telephone him during certain hours, he became accustomed to surprise events during his performances, and learned how to ignore them.

It would be difficult to argue that the reasoning done by this pianist was not planning—it is indeed an example of the use of a strategy for learning to ignore distractions during execution. This same strategy might also be useful for surgeons and soldiers, for whom distractions can be deadly. While this strategy may be abstractly defined in such a way that it could be applied across all goal-directed domains, it cannot be done so without carrying along with it the set of modeling assumptions that it entails. Along with the general planning concepts of operators, conditional effects, and preconditions, this strategy has at its core the more subtle planning concepts of *execution failures*, *distracting events*, *attention*, *ability to ignore*, and *practice*, among others. Some existing formal commonsense theories include predicates or other notation that refer to some of these concepts, while other concepts have not yet received formal consideration. What we are arguing for here, however, is that any complete representational theory of commonsense reasoning about planning will include a mechanism for expressing all of them. In analyzing this one strategy and identifying the concepts that must play a role in its definition, we have gathered some weak evidence for the representational requirements of such a theory.

The opportunity here stems from the fact that weak evidence can be aggregated. By systematically analyzing everyday planning strategies on a large scale, and by authoring hundreds of strategy definitions, we can aggregate the evidence provided by each individual strategy to identify the scope and scale of the representation problem and the array of specific planning concepts that must be expressible in a complete, formal representational theory. The remaining sections of this article describe a project to pursue this approach, one that is composed of three parts. In the first part of the project, hundreds of strategies were collected from a wide range of different real-world planning domains. In the second part, representations of each of these strategies were authored so as to identify the specific conceptual assumptions that they make about the human planning process. In the third part of this project, all of the concepts used in these representations were extracted from their strategy contexts, and organized into sets of component theories that collectively describe representational elements of our commonsense models of the human planning process.

As a research methodology, the approach that is described in this article offers an alternative to the logicians' agenda for the exploration of representation in commonsense reasoning that is not based on the competency of theories to mirror human inferences. As such, the products of this approach have no inferential capacity on their own—the individual concepts that are identified through this methodology outline only the representational elements that would compose axiomatic theories to drive commonsense deductive inference. The specific value of this methodology to logicians whose primary interest is deductive inference is in the lists of representational terms, which suggest specific concepts that should be expressible in future axiomatic theories. While many of the representational areas identified in this project have received a substantial amount of formal attention, few existing formalizations have complete coverage over all of the terms in any given area. Furthermore, there remain several representational areas that have not yet been the subject of formalization efforts. For Artificial Intelligence researchers whose primary concern is not deductive inference (e.g., natural language processing), the opportunity presented by this approach is to treat the problem of commonsense representation as separable from commonsense inference, and to capitalize on progress in the former while the research community makes comparably slower progress on the latter.

2. Collecting strategies from real-world domains

In selecting an appropriate corpus of strategies to serve as the subject of this analysis, the primary characteristic that is desired is breadth. The ideal collection would span across the widest range of planned human behavior, and capture the commonsense strategies that people use and refer to in all aspects of their lives, from the activities of personal life to those of occupational pursuits. Breadth in the strategy collection helps avoid a pitfall of any analytical approach: the problem of generalizing over a sample that is not representative of the population. Whereas most scientific fields turn to random sampling of a population to avoid this problem, it isn't clear exactly what this would mean in the analysis of strategies. If there already existed an exhaustive catalog of commonsense strategies, perhaps weighted by the frequency that they are referenced across the worlds' cultures, then this list could be sampled to assemble a representative set for analysis. As no such catalog exists, the first part of our efforts was to assemble a collection to use, with the characteristic of breadth as the primary aim.

One factor that greatly aided this collection process is the association between strategies and specific planning domains. While strategies are, in general, applicable across multiple domains, the phrases that people use to refer to strategies often are grounded in specific areas of pursuit. For example, the strategy of employing the spies of your enemy as double agents is easiest to reference in the language of international espionage (as it is done here), but the underlying strategy does not require that governments are the planning agents, that the spies are professionals, or that the information they obtain is political or military in nature. The same strategy may be equally useful in foiling the attempts of a sports team that is attempting to spy on the tactics and plays of their rival as it is in learning about adversarial countries. Yet this strategy seems to find a natural home within the domain of international espionage, presumably because the important actions, people, places, and things of this strategy are all clearly defined, easily referenced, and part of a broader commonsense understanding of this domain. The affinity of strategies to certain domains enables a domain-driven approach to their collection. By selecting first a set of domains, several different techniques can be employed to identify a collection of strategies that are associated with that domain in some way.

A first technique that can be used is to collect strategies that are referenced in existing texts that are nearly encyclopedic of strategies in their respective domains. Examples of these sorts of texts include Sun Tzu's *The Art of War* and Niccolo Machiavelli's *The Prince*, which describe a broad range of strategies in the domains of warfare and governance, respectively. What set these works apart from other non-fiction writing in a domain are three important features. First is the sheer quantity of unique strategies that are discussed; both of these texts, for example, explicitly discuss several dozen strategies. Second is the extremely high ratio of text in these works that focuses on these strategies to that which doesn't. Third is the degree to which the strategies that they contain have pervaded cultures throughout the world over time, such that few of the comments made by their authors strike the contemporary reader as particularly innovative—indeed, the strategies they discuss are now simply commonsense.

A second technique that can be used to collect strategies from a domain is directed interviewing of practitioners in the domain. While it is clear that experts in any given field (e.g., marketing strategists, politicians, musicians) use knowledge that is highly specific to their specialty, they tend to find ways of explaining their expertise to novices that rely more heavily on the commonsense knowledge that is shared between them. When interviewers who are domain novices engage domain experts in a conversation about the strategies that they employ, many experts will choose to base their explanations on the strategies that are naively attributed to their domain. The attentive interviewer can listen specifically for these references (sometimes ignoring the actual explanations) in order to collect the set of associated strategies.

A third technique for collecting strategies is to observe and interpret the planning and problem solving of other people as they are engaged in tasks within a domain. While more scientific observations of behavior strive to objectively gather data about the actual performance of subjects, the intent of interpretive observation is to gather the strategies that are subjectively recognized and reported by the *observers* as they are exposed to the behavior of other people. Interpretive observation of this sort may be most effective when the researcher describing the strategies is also a practitioner within the domain, allowing the researcher to make note of strategic planning behavior within the domain over extended periods of time.

While these three techniques lack many of the qualities that investigative researchers would like to have of their methods, they serve as adequate precursors to the more rigorous and repeatable methods that need to be developed as research in commonsense reasoning progresses. Each of these three techniques was employed to collect the strategies used in this investigation. The domains that were investigated were determined by the availability of the resources necessary to utilize these three techniques, namely access to texts that were encyclopedic of strategies within a domain or access to domain practitioners that could participate in directed interviews or interpretive observation. In all, ten different planning domains were explored. From each of these domains, dozens of strategies were collected, yielding a corpus of strategies consisting of 372 entries. Each strategy entry consisted of a few descriptive sentences and a short title so that it could be easily referenced. The ten domains are described briefly here, along with the number of strategies collected from each one and an example entry:

(1) Animal Behavior (37 strategies). An analysis of an introductory zoology text revealed the strategies that people attribute to animals to explain their behavior in an anthropomorphic manner. Example: *Cowbirds and Cuckoo birds abandon their eggs in the nests of other bird species that will raise the hatchlings along with their own offspring.*
(2) Business (62 strategies). Directed interviews with students of business administration and investment law produced a collection of strategies used in the business areas of marketing, production, and acquisitions. Example: *Companies that compete for repeat business from consumers gain customer loyalty by issuing redeemable points (e.g., frequent flyer miles) with each purchase.*
(3) Counting (20 strategies). Interpretive observation was used to identify the ways that people perform various object counting tasks, where variation is determined by characteristics of the situation. Example: *When counting identical things organized in a circle, people will mark their starting location to avoid over-counting.*
(4) Education (32 strategies). Directed interviews with former primary and secondary school teachers produced a set of strategies that focused on assessment, teaching methods, and techniques for motivating students. Example: *Occasionally delivering an unscheduled "pop quiz" encourages students to study materials in the duration between larger exams.*
(5) Immunology (21 strategies). An analysis of several introductory immunology textbooks produced a set of strategies that anthropomorphically describe the behavior of cells and viruses as goal-directed agents. Example: *One way that cells guard against the spread of viruses is to destroy themselves through programmed cell death when an infection is detected.*
(6) Machiavelli (60 strategies). Niccolo Machiavelli's sixteenth-century text *The Prince* was analyzed to collect a set of strategies to be used by government leaders to obtain and maintain their power. Example: *If established as a lord by the nobility of a territory, a new prince should work to become popular to establish a power base among the citizenry.*
(7) Performance (39 strategies). Directed interviews with concert pianists were used to collect the strategies that musicians use to manage their performance abilities. Example: *To avoid being derailed by distracting events during a performance, musicians will purposefully cause surprising distractions to occur during their practice sessions in order to become accustomed to them.*
(8) Relationships (22 strategies). Interpretive observation of dating couples was used to identify a set of strategies that people use to find, keep, and leave a loving partner. Example: *On a first date, a skeptical person may ask a friend to telephone halfway through to give them a chance to fabricate an emergency that requires them to leave.*
(9) Science (34 strategies). Directed interviews and interpretive observation were used to collect a set of strategies that research scientists use to advance their fields, pursue

funding, and garner attention to their work. Example: *To moderate the difficulty of obtaining research funds, many researchers tie their work to some application domain that is backed by wealthy interests, such as the military or pharmaceutical companies.*

(10) Warfare (45 strategies). *The Art of War*, written by Chinese military strategist Sun Tzu around the fifth century BC, was analyzed to collect strategies pertaining to waging war against military adversaries. Example: *By leaving the enemy a means of retreat, the attacking army prevents the threat of extraordinary heroic defenses on the part of a desperate adversary.*

A preliminary investigation of planning behavior in the areas of board games (backgammon) and athletic sports (basketball) indicated that these areas would also be promising sources of strategies, particularly given the abundance of texts that are nearly encyclopedic of tactics for any particular type of game. However, in beginning to collect strategies from these areas it became evident that many of the strategies discussed were geared toward exploiting opportunities that exist within the constraints of the rules of a specific game. While in some cases these strategies were applicable across multiple game genres, strategies for exploiting game rules did not appear to have the same degree of cross-domain applicability as those collected from other domains. As this initial investigation of planning strategies was aimed at identifying representations that would be pervasive across planning tasks, the decision was made to focus attention on the 10 planning domains just listed, and to leave the study of game strategy for future research.

3. Pre-formal strategy representations

One of the first attempts to explore the computational role of strategies in human planning and problem solving was in the Computer Science dissertation of Gregg Collins at Yale University [2]. In this work, Collins identifies a set of two dozen strategies and describes how each would be applied in the competitive domain of American football. Many of these strategies could be referenced with common phrases (e.g., “hold the fort”, “beat them to the punch”, and “hedge your bets”). The importance of Collins work was to recognize that each of these strategies could be defined by identifying its component parts. To decompose each of these strategies, Collins listed a set of statements about the world that would trigger the use of the strategy, along with a set of directives for the planning system.

For example, consider a strategy that Collins labeled “best available resource”. The strategy is one that would be applied when a planner is devising a plan for acquiring resources over time, as the coach of a professional sports team would use to select new athletes from all of those that were available. While some sports teams acquire new players that best meet a current need, this strategy argues for selecting the best player available, regardless of whether they would best serve the role that is needed. For example, an American football coach might select an outstanding wide receiver player over a fairly good quarterback, even if it is the role of quarterback that is lacking on the current team roster. Collins offers a decomposition of this strategy as follows:

Best Available Resource:

Given: A cyclical opportunity to acquire a resource.
Some diversity in the specific roles played by resource items.
Some way of comparing items across roles.
A distribution of available resources that matches the distribution of needs over time.
An acquisition plan based on filling roles as they become needed.

Produce: A new acquisition plan in which the best available resource items are always selected, regardless of any analysis of immediate needs.

While there may be an interesting debate to be had regarding the inclusion of these particular statements in the definition of this strategy, Collins succeeds in showing that strategies are compositional in nature, where components can be defined using terms that are not specific to any one domain. In describing the strategy in abstract terms, Collins generalizes the definition of this strategy to one that covers a set of planning cases that are all analogous to this one case of selecting athletes for a sports team. As Collins notes, the same approach could be used over time by someone shopping for clothes, where desirable clothes that are on sale are preferred to clothes that fill a need in the collection of clothes that the shopper already possesses. There are few surface-level similarities between the domains of shopping for clothes and sports management, but Collins' definition of this strategy captures the essential common features of these two cases that cause us to recognize them as instances of exactly the same strategy.

As a knowledge structure to serve in computational models of cognition, this definition suffers greatest in its ambiguity. By writing this definition as a collection of English phrases instead of as a formal representation (i.e., logic statements, frame representation, or semantic net), its computational utility is greatly constrained. In its current form, it can be used neither to support the planning process when faced with a resource allocation problem, nor to make inferences about other people that are employing this strategy. To enable the possibility of using this strategy in these types of computation, it must be expressed in a formal manner. That is, it must be expressed in a language containing a set of controlled vocabulary terms (or symbols) coupled with a clearly defined syntax.

The manner in which Collins defined these strategies should not be viewed as an oversight. Indeed, Collins' definitions reflect a problem in the knowledge representation community that persists today. Simply put, no formal language has yet been developed that has a vocabulary that is broad enough to fully represent these strategies. The pursuit of such a language has yielded steady progress, both in the field of Artificial Intelligence planning (e.g., [7]) and in the area of commonsense reasoning about human planning (e.g., [11]), but these current efforts fall short in their ability to capture many of the subtleties that Collins refers to in the strategies that he defines. In the example given above, some of the problematic concepts include that of distributions of needs and available resources, cyclic opportunities, comparisons of items across roles, and analysis of immediate needs.

While the current lack of a rich formal language prohibits the expression of this strategy in a completely formal manner, two improvements on Collins' representations can be made to move them in the right direction. First, each of the references to specific concepts (such as the problematic list above) can be tagged in the text, marking it as a term that will be part of a complete formal language, or as a concept that must be further defined by those terms that are. Second, each of the English phrases that compose the definition can be written in a manner that reduces the use of these composite terms, favoring references to concepts that are believed to be non-reducible, or perhaps reducible in obvious ways. In applying these improvements to the example strategy above, a concept like *cyclic opportunities* could be tagged as a required part of some future formal language, although it may be more appropriate to change the definition such that it references two component parts of this concept, *opportunities* and *recurring events*.

The effect of these improvements is to produce definitions that are not entirely informal or formal, but somewhere in between. They are perhaps best described as *pre-formal representations*, in that they refer to representational elements that will have to be included in future formal representation languages. The aim of pre-formal representation is to make strong commitments to the inclusion of certain concepts in a representational vocabulary without making strong commitments toward selecting a specific lexical referent (or symbol) or their proper syntactic form. The "pre" in pre-formal is forward-looking to a time when our representation languages contain all of the notations necessary (perhaps as predicates in a first order predicate calculus) to express these definitions in a satisfactory formal manner.

This pre-formal approach, which improves upon Collins' representational forms, was used to represent the strategies described in this article. In using a pre-formal approach to create representations of the 372 strategies that were collected, the aim was to author exact definitions of these mental structures with the least amount of ambiguity. As in Collins' strategy definitions, the components of pre-formal representations of strategies generalize the definition to cover the full set of planning cases that are analogous to the one collected from a specific domain. The aim is to describe every planning situation that is an instance of the strategy (regardless of the domain in which it appears), and no others. Unlike the representations of Collins, which group definition phrases into "given" and "produce" categories, the definitions in the current work are each authored as a set of grammatical English sentences in the present tense. The person who is employing the strategy is referred to as "the planner" and others that play roles are "agents". In order to mark certain words or phrases as terms that will need to be part of any future formal language to represent these strategies, they are capitalized and italicized in the text of the pre-formal representation.

Three examples of the 372 pre-formal strategy representations are presented next. Among the entire set of strategy representations, a total of 8,844 words and phrases were tagged as concepts that will need to be expressible in a complete formal vocabulary, an average of 24 concepts for each pre-formal representation.

> Business strategy 59, Macaroni defense: Prevent takeover by issuing bonds to be redeemed at high cost if a takeover occurs.
>
> The planner has a *Leadership role* in an *Organization*, and *Envisions a threat* to goals of the *Organization* that a different agent A1 will execute a plan that includes *Trade* of an *Amount of resources* of the agent A1 for *Portions of ownership* that are a *Majority portion*. The planner executes a plan to *Make an offer* to a different agent A2 where

the *Offer requirement* is that the agent A2 *Transfers resources* to the *Organization* of an *Amount of resources*, and the *Offered action* is that the planner will *Schedule* the *Transfer of resources* to the agent A2 of an *Amount of resources* that is greater than in the *Offer requirement*, with the *Offer condition* that *If it is the case* that if an *Arbitrary agent* executes the plan that is the plan in the *Threat*, the *Organization* will *Enable* the agent A2 in the offer to cause the *Organization* to *Reschedule* the *Offer action* with an *Amount of resources* that is greater. The planner then *Schedules a plan* to *Inform* the agent A1 in the *Threat* of this offer after the *Successful execution* of the plan.

Machiavelli strategy 48, Foster then crush animosity to improve renown: "For this reason many consider that a wise prince, when he has the opportunity, ought with craft to foster some animosity against himself, so that, having crushed it, his renown may rise higher."

Representation: The planner has a *Role* in a *Power relationship* over a *Community* and has the goal that a set of agents A1 *Believe* that the planner is the *Best candidate* for this role. There exists an agent A2 that has goals that are in *Goal conflict* with the planner, and that has had a *Planning failure* to achieve their goals with a *Cause* of *Lack of resources* or *Lack of agency*. The planner plans and executes for the goal that the agent A2 *Believes* that there exists an *Opportunity* for this agent A2 to achieve their goals, where no *Opportunity* exists. The planner then *Monitors* the agent A2 for the *Start of execution* of plans to achieve the goals. When this occurs, the planner executes a plan to *Destroy* the agent A2, with the goal that the first set of agents A1 have *Event knowledge* of the planner's execution.

Warfare strategy 44, Give false information to enemy spies: Secretly use enemy spies to send deceptive information to your enemy.

Representation: The planner has an *Adversarial relationship* with another agent A1. An agent A2 has a *Cooperative relationship* with the agent A1 to execute *Cooperative plans* that include *Informing* the agent A1 of *Information* that involves the planner. The agent A2 has the *False belief* that the planner has a *False belief* that the planner has a *Cooperative relationship* with the agent A2. The planner *Monitors* planning for *Adversarial plans* that have a *Threat* that the adversary agent A1 will *Successfully execute* a *Counterplan*. *In this case*, the planner executes a plan to cause the agent A2 to have a *False belief* about the *Adversarial plan*, and then *Enables* the *Cooperative plan* of this agent A2 that includes *Informing*.

4. Recurring strategy patterns

In electing not to use a predetermined strategy frame in these pre-formal representations, such as the "given" and "produce" frame seen in the work of Collins, it was possible to take a somewhat more empirical approach to the determination of such a frame. If there were any inherent structures in strategies, they would possibly appear as recurring patterns in the pre-formal style of representation that was used here. The finding was that, while no single generalization universally characterizes all strategies, there were a number of recurring structural features. In general, a frame was evident that separates a strategy into four components, corresponding to descriptions of a situation, a goal, an envisioned possibility, and a set of planning directives that may lead to the achievement of the goal.

The first two components that are evident in these strategy representations describe a state of the world where the planning agent might consider the application of the strategy. First, there is a description of the situation, which often refers to the types of relationships that the planner has with other people, the state of various plans and processes, and the actions that the planner or other people have already undertaken. Second, within the scope of the situation is a description of the goal of the planner, an apparently inherent structural component that all representations contained. If strategies are viewed as rules in a rule-based planning system, then it is these first two components that serve as the triggering conditions for further consideration of a strategy.

This further consideration is an active reasoning process, which is reflected in a third structural component that appears often, but not always. Here, the planning system determines if the strategy could be applicable to meet its goal by imagining something that isn't already a part of their knowledge—a possible future event, a consequence of an action that hasn't yet been taken that poses a threat or an opportunity, or an imagined world where some specific feature has changed. While it may be simpler to consider these envisioned possibilities as an additional part of a strategy's triggering conditions, they imply an active process of *envisionment* (as in [4]) as opposed to the more straightforward pattern matching that could be used to recognize an applicable situation and goal.

The strategy representations universally concluded with a set of planning directives that would move the planner closer to the goal. Sometimes these were simply abstract plans in the traditional sense, that is, a partially ordered set of actions whose execution would cause the attainment of the goal. More often, actions were combined with more algorithmic control instructions, such as a directive to adapt an existing plan in a particular way, to begin the process of monitoring for a specific event, or even to modify the planner's own view of what their goal actually is. In many cases, the actual plans to be executed were left completely unspecified, with directives that initiate the planning process itself for a newly added goal, possibly with the addition of certain planning and scheduling preferences or constraints.

In contrast to these four structural components that were generally included in strategy representations, little attempt was made to include a complete causal account of how the execution of a given strategy achieves the planner's goal. That is, the explanation of how a particular strategy works was not generally included in its representation. For example, the business strategy "Macaroni defense" described in the previous section includes a triggering situation, a goal of the planner, and a set of planning directives, but the causal chain that would lead from the execution of the planning directives to the achievement of the goal remains implicit. In the business case, it can be inferred that informing the acquiring company that the target company has outstanding bonds that must be paid in the event of a takeover will cause the acquiring company to abandon the takeover plan, as the target company's value would be undermined. The expectation is that rational planning agents would need to reason through this chain of causality in order to fully evaluate the appropriateness of this strategy for a given planning situation. The convention used in this project was to not account for this sort of causal reasoning in the representations of the strategies themselves, but to suggest that causal explanations of this sort might be generated when applying a strategy to a particular planning problem. Part of the rationale for this decision stems from the central goal in authoring exact strategy definitions, where the component representations describe every planning case that is an instance of a strategy and no others. By not including representations of the causal effects of strategies, the claim is made that this information does not further discriminate between positive and negative examples of strategy instances.

5. Organizing representation terms

By tagging significant words and phrases in these pre-formal representations (with italics and capitalization), it was possible to extract from this set a list of references to all of the concepts that must be expressible in any complete formal language for strategic planning. To best make use of this list, an effort was made to turn this list into a controlled vocabulary—where a single term represents all of the synonymous occurrences of any given concept in the set of strategy representations. The resulting controlled vocabulary can then be further organized in a manner that best informs the development of a future formal language.

In the 372 representations, there were 8,844 instances where words and phrases were tagged as significant concepts. To develop a controlled vocabulary from this list, three steps were necessary. First, duplicate occurrences of any given lexical usage were removed. Second, a single lexical form was chosen for each set of forms that differed only in their inflection, combining multiple verb tenses or plural and singular forms into one. Third, the remaining forms were organized into groups of synonymous referents, where every member was associated with a single concept, from which one was selected as the controlled term for the concept. Of these three steps, it is this last step that most requires the consideration of the author of the original pre-formal representations. In most cases, the grouping of synonymous terms is an easy one; for example, the notion of a "*Potential motion*" is synonymous to a "*Possible motion*". However, the decision of whether "*Free time*" is synonymous with an "*Unscheduled duration*" is one that requires a review of how these terms were used in the representation contexts in which they appear.

By controlling the vocabulary in this manner, the initial list of lexical forms was reduced to a vocabulary of 988 unique concepts. To better understand this list, an analysis was performed to determine the frequency in which each term was used within the 372 representations. In ordering the frequencies for each term from most used to least used, it was evident that a small percentage of terms were ubiquitous throughout the 372 representations, but the majority of terms were used only a handful of times. The rate of growth of new terms decreased over the course of the authoring process with the addition of each new representation, but remained positive at the end of this work.

With close to one thousand concepts in the vocabulary, an attempt was made to understand its scope by sorting and organizing this list into groups of associated terms. When the final list was compiled, it was evident that many of the terms referred to concepts that had previously been formalized as part of existing representational theories, or that are related to processes that have been the subject of cognitive modeling efforts in the past. Using these theories and processes to designate different groups, the 988 terms were organized into 48 representational areas, each of which is described briefly in this section.

The intent of presenting this list of representational areas here is to allow us to evaluate the current state of the field of knowledge representation as it relates reasoning about intentional behavior. Many of the first several representational areas in this list have previously received a great deal of attention in the field of commonsense reasoning. These areas, which include representations of time, space, and events, overlap with the breadth of representational areas that have been formalized to build models of qualitative reasoning in physical domains. In contrast, many areas presented here specifically support reasoning about the mental life of people—their goals, plans, and cognitive abilities—which have received comparably little attention as component representational theories. For these areas, this list aims at defining the scope of the research problem at hand as a set of component representational theories that will need to be developed into syntactically formal notations paired with appropriate inference mechanisms.

In reviewing the 48 representational areas that organize the terms used for strategy representation, some generalizations about their scope can be made. The first several areas (listings (1)–(8)) are very similar to the sorts of representational theories that have been identified as fundamental to commonsense reasoning about physical systems (e.g., in [8]). The second several areas (listings (9)–(22)) introduce intentional agents, their plans and goals, and the roles they play in relationships with other agents. This second set is receiving increasingly more formal attention within the commonsense reasoning research community as these classes of representations are seen as critical to certain human-computer interaction technologies. The last half of this list (listings (23)–(48)) describe the mental processes of people, including capacities for envisioning states, selecting plans, and mechanisms for executing plans in pursuit of their goals. While these areas have always been the interest of researchers that build computational models of cognitive processes, they have received very little attention as representational theories.

The 48 representational areas are as follows:

(1) Time (23 terms): Moments in time, including the current moment, compose durations that have ordering relationships with other moments and durations, and may have indefinite start times or end times. Examples: *Current moment*, *Immediately before*, *Variable duration*.
(2) States (18 terms): The world is described in states that change, are partially observable, and where portions of these descriptions define environment contexts. Examples: *Future state*, *Unchanged state*, *Persistent state*.
(3) Events (11 terms): Events are interpretations of states changing over time, and may occur multiple times with some frequency, and have temporal relationships with other events. Examples: *Concurrent event*, *Frequency*, *Periodic event*.
(4) Space (40 terms): Space is composed of locations in regions with boundary definitions, where locations can define paths through space with directionality and endpoints. Examples: *Boundary definition*, *Linear path*, *Direction*.
(5) Physical entities (24 terms): Physical objects and substances, including the physical bodies of agents, are composed of components, can have some state of configuration,

and can contain or be attached to other things. Examples: *Substance*, *Contained thing*, *Configurable state*.

(6) Values and quantities (21 terms): Qualitative values and their quantifications may be defined by their ranges or in relationship to others, and can describe amounts in the world. Examples: *Variable amount*, *Unknown quantity*, *Directly proportional*.

(7) Classes and instances (21 terms): Things in the world are conceptualized as instances of class categories at various levels of abstraction, with characteristics that are sometimes specified or unknown. Examples: *Class instance*, *Unknown class*, *Characteristic change*.

(8) Sets (31 terms): Things in the world can be conceptualized as being members of sets that have certain membership rules, and which may be ordered by applying some ordering principle. Examples: *Arbitrary member*, *Minority portion*, *Next in order*.

(9) Agents (22 terms): Agents are intentional things that have certain characteristics and roles, and are distinguished by their relationships to other agents. Examples: *Adversarial agent*, *Agent role*, *Unknown agent*.

(10) Agent relationships (18 terms): Relationships between agents are generally distinguished by who does the planning, who does the execution, and whose goals are in pursuit. Examples: *Competitive relationship*, *Persistent relationship*, *Symmetric relationship*.

(11) Communities and organizations (10 terms): Sets of agents may be described as organizations and communities, where agents have roles to play determined by the structure of these sets, and where collective processes and goals may exist. Examples: *Organizational goal*, *Top organizational role*, *Organizational agency*.

(12) Goals (27 terms): Goals of agents describe world states and events that are desired, and includes both states and events that are external to the planner as well as those that characterize desired internal mental states and processes. Examples: *Auxiliary goal*, *Knowledge goal*, *Shared goal*.

(13) Goal themes (6 terms): A potential reason that an agent may have a goal can be based on the role that the agent has in relationships and organizations, or because of a value that they hold. Examples: *Generous theme*, G*ood person theme*, *Retaliation theme*.

(14) Plans (30 terms): The plans of agents are descriptions of behaviors that are imagined to achieve goals, and can be distinguished by the types of goals that they achieve or by how they are executed, and may be composed of other plans or only partially specified. Examples: *Adversarial plan*, *Repetitive plan, Shared plan*.

(15) Plan elements (30 terms): Plans are composed of subplans, including branches that are contingent on factors only known at the time of execution. They may have iterative or repetitive components, or include components that are absolutely required for a plan to succeed. Examples: *If/then*, *Iteration termination condition*, *Triggered start time*.

(16) Resources (16 terms): Classes of things that are reusable across a broad range of plans are resources, which are described in terms of their amount, and can be managed in a variety of ways. Examples: *Expended resource*, *Resource class*, *Transfer resource*.

(17) Abilities (13 terms): Ability is a characteristic of an agent that is predictive of the successfulness of their execution of a plan, and may be qualified at various degrees of skill. Examples: *Ability level*, *Competent ability*, *Unknown ability*.

(18) Activities (16 terms): Activities are shared plans and expectations about behavior where agents have roles and where events follow an expected script. Activities are also distinguished by whether they are part of an agent's set of normal behaviors. Examples: *Activity duration*, *Agent with role*, *Normal activities*.
(19) Communication acts (23 terms): Some abstract plans deal with transferring information to other agents, and include making and accepting offers, asking and answering questions, requesting permission, persuading and threatening. Examples: *Accept permission request*, *Offer condition*, *Threatened action*.
(20) Information acts (12 terms): Some types of beliefs, which may be true or false, can be externalized as information, sometimes encoded into some external physical thing or passed as signals. Examples: *False information*, *Observed signal*, *Decode information*.
(21) Agent interaction (21 terms): Many recognizable abstract patterns of events involve the interaction of agents, such as the assignment of ownership over a thing, a defense against an attack, and the execution of offered actions. Examples: *Accept candidate*, *Assist*, *Trade*.
(22) Physical interaction (13 terms): Many recognizable abstract patterns of events involve the interaction between agents and physical things, such as configuring a thing into a particular state and the use of something for its designed purpose. Examples: *Configure thing*, *Copy thing*, *Instrumental use*.
(23) Managing knowledge (30 terms): The knowledge that agents have is a set of beliefs that may be true or false based on certain justifications, and can be actively assumed true, affirmed, or disregarded entirely. Examples: *Assumption*, *Justification*, *Revealed false belief*.
(24) Similarity comparison (16 terms): Agents can reason about the similarity of different things using different similarity metrics, where analogies are similar only at an abstract level. Examples: *Class similarity*, *Similarity metric*, *Make analogy*.
(25) Memory retrieval (3 terms): Agents have a memory that they use to store information through a process of memorization, and may use memory aids and cues to facilitate retrieval. Examples: *Memory cue*, *Memory retrieval*, *Memorize*.
(26) Emotions (8 terms): Agents may experience a wide range of emotional responses based on their appraisal of situations, which defines their emotional state. Examples: *Anxiety emotion*, *Pride emotion*, *Emotion state*.
(27) Explanations (17 terms): Agents generate candidate explanations for causes in the world that are unknown, where preferences for certain classes of explanations are preferred. Examples: *Candidate explanation*, *Explanation preference*, *Explanation failure*.
(28) World envisionment (48 terms): Agents have the capacity to imagine states other than the current state, to predict what will happen next or what has happened in the past, and to determine the feasibility of certain state transitions. Examples: *Causal chain*, *Envisioned event*, *Possible envisioned state*.
(29) Execution envisionment (23 terms): One mode of envisionment is that of imagining the execution of a plan for the purpose of predicting possible conflicts, execution failures, side effects, and the likelihood of successful execution. Examples: *Envisioned failure*, *Side effect*, *Imagine possible execution*.

(30) Causes of failure (31 terms): In attempting to explain failures of plans and reasoning, agents may employ a number of explanation patterns, such as explaining a scheduling failure by the lack of time, or a planning failure by a lack of resources. Examples: *False triggered monitor*, *Lack of ability*, *Successful execution of opposing competitive plan*.

(31) Managing expectations (8 terms): Envisionments about what will happen next constitute expectations, which can be validated or violated based on what actually occurs. Examples: *Expectation violation*, *Unexpected event*, *Remove expectation*.

(32) Other agent reasoning (8 terms): Envisionments about the planning and reasoning processes of other agents allow an agent to imagine what he would be thinking about if he were the other agent. Examples: *Guess expectation*, *Guess goal*, *Deduce other agent plan*.

(33) Threat detection (15 terms): By monitoring their own envisionments for states that violate goals, an agent can detect threats and track their realization. Examples: *Envisioned threat*, *Realized threat*, *Threat condition*.

(34) Goal management (27 terms): Agents actively manage the goals that they have, deciding when to add new goals, commence or suspend the pursuit of goals, modify or specify their goals in some way, or abandon them altogether. Examples: *Currently pursued goal*, *Goal prioritization*, *Suspend goal*.

(35) Planning modalities (17 terms): The selection of plans can be done in a variety of different ways, such as adapting old plans to current situations, collaboratively planning with other agents, and counterplanning against the envisioned plans of adversaries. Examples: *Adversarial planning*, *Auxiliary goal pursuit*, *Imagined world planning*.

(36) Planning goals (28 terms): The planning process is directed by abstract planning goals of an agent, which include goals of blocking threats, delaying events, enabling an action, preserving a precondition, or satisfying the goals of others. Examples: *Avoid action*, *Delay duration end*, *Maximize value*.

(37) Plan construction (30 terms): Agents construct new plans by specializing partial plans, adding and ordering subplans, and resolving planning problems when they arise. Examples: *Candidate plan*, *Planning failure*, *Planning preference*.

(38) Plan adaptation (18 terms): Existing plans can be adapted and modified by substituting values or agency, and by adding or removing subplans to achieve goals given the current situation. Examples: *Adaptation cost*, *Adaptation failure*, *Substitution adaptation*.

(39) Design (8 terms): One modality of planning is design, where the constructed plan is a description of a thing in the world within certain design constraints, and where the resulting things have a degree of adherence to this design. Examples: *Design adherence*, *Design failure*, *Designed use*.

(40) Decisions (38 terms): Agents are faced with choices that may have an effect on their goals, and must decide among options based on some selection criteria or by evaluating the envisioned consequences. Examples: *Best candidate*, *Decision justification*, *Preference*.

(41) Scheduling (22 terms): As agents select plans, they must be scheduled so that they are performed before deadlines and abide by other scheduling constraints. Plans may

have scheduled start times and durations, or may be pending as the planner waits for the next opportunity for execution. Examples: *Deadline*, *Pending plan*, *Scheduling constraint*.

(42) Monitoring (18 terms): Agents monitor both states and events in the world and in their own reasoning processes for certain trigger conditions which may prompt the execution of a triggered action. Examples: *First monitor triggering*, *Monitoring duration*, *Monitor envisionment*.

(43) Execution modalities (11 terms): Plans can be executed in a variety of ways, including consecutively along with other plans, in a repetitive manner, and collaboratively along with other agents. Examples: *Concurrent execution*, *Continuous execution*, *Periodic execution*.

(44) Execution control (29 terms): A planner actively decides to begin the execution of a plan, and may then decide to suspend or terminate its execution. A suspended plan can later be resumed from the point at which the agent left off. Examples: *Execution delay*, *Suspend execution*, *Terminate activity*.

(45) Repetitive execution (16 terms): Some plans and subplans are executed iteratively for some number of times, or repetitively until some termination condition is achieved. Examples: *Current iteration*, *Iteration completion*, *Remaining repetition*.

(46) Plan following (30 terms): Agents track the progress of their plans as they execute them in order to recognize when deadlines are missed, preconditions are satisfied, and when they have successfully achieved the goal. Examples: *Achieve precondition*, *Miss deadline*, *Successfully executed plan*.

(47) Observation of execution (28 terms): Agents can track the execution of plans by other agents, evaluating the degree to which these executions adhere to performance descriptions known to the observing agent. Examples: *Observed execution*, *Assessment criteria*, *Performance encoding*.

(48) Body interaction (15 terms): The physical body of an agent translates intended actions into physical movements, and sometimes behaves in unintended ways. The body modifies the planner's knowledge through perception of the world around it, and by causing a sensation of execution. Examples: *Impaired agency*, *Nonconscious execution*, *Attend*.

6. Conclusions

A common criticism of artificial intelligence research in general is that proposed solutions often do not scale to the degree necessary to address large, real-world problems. In developing formal theories of commonsense knowledge, the problem is compounded by our uncertainty about the scale that would be required. That is, there are few methodologies available to researchers in this area that can be used to define the scope of the problem. Researchers interested in qualitative reasoning about physical systems have, in the past, proposed a number of long-term formalization efforts with only a vague sense of the scope of the work that remains [3,8]. The methodology of large-scale pre-formal representation that is described here can be seen as a tool to estimate the true scope of required representational theories to support commonsense reasoning.

The application of this methodology as it is described here has defined the scope of the representational requirements for reasoning in intentional domains—reasoning about the situations, plans, goals, and actions of intentional agents. The key to the success of the application of this methodology was the breadth of the subject of analysis, strategies. While other types of knowledge structures, such as narratives and explanations, may cover much of the same representational territory, the relative ease with which strategies could be collected and distinguished from each other greatly facilitated this approach. If this methodology is ever applied to other areas of commonsense reasoning (such as reasoning about physical systems), the first challenge would be to identify an appropriate knowledge structure of equal breadth within the area that also shares these characteristics.

This work has explored the utility of pre-formal representation for large-scale knowledge engineering, where definitions can be quickly authored as a set of tagged natural-language sentences. These pre-formal representations allow a knowledge engineer to make strong commitments concerning the semantic content of a representation without committing to a particular syntactic notation. The primary utility of the use of pre-formal representations in this work was to gather evidence for the concepts that will need to be expressible in future formal languages. However, an additional benefit to using this style of representation was that it enabled an authoring effort that is much larger in scale than typically seen in representation projects. In focusing on scale rather than syntactic formality, an understanding of the breadth of representational requirements was achieved with a tractable amount of knowledge representation effort.

This research has focused primarily on the concepts that must be expressible in future formalizations, but full formal theories of intentional action will also require the inclusion of all of the inference mechanisms that traditionally form the basis of automated commonsense reasoning systems. While many of the representational areas that were identified here have already received some amount of formal attention, there remain many areas that have not. The concepts that are grouped into each of these areas may provide a basis for evaluating the conceptual coverage of any proposed formal language, but the majority of the remaining work will involve authoring sets of axioms for each of these areas to drive the automated reasoning process. Representing strategies on a large scale has identified the scope of the remaining formalization work, providing investigators a roadmap for future commonsense reasoning research.

References

[1] P. Cohen, H. Levesque, Intention is choice with commitment, Artificial Intelligence 42 (3) (1990) 213–261.

[2] G. Collins, Plan creation: Using strategies as blueprints, Doctoral Dissertation, Yale University, New Haven, CT, 1987.

[3] E. Davis, The naive physics perplex, AI Magazine 19 (4) (1998) 51–79.

[4] J. de Kleer, J. Brown, A qualitative physics based on confluences, Artificial Intelligence 24 (1984) 7–83.

[5] G. Ferguson, Knowledge representation and reasoning for mixed initiative planning, Doctoral Dissertation, Computer Science Department, University of Rochester, Rochester, NY, 1995.

[6] R. Fikes, N. Nilsson, STRIPS: A new approach to the application of theorem proving to problem solving, Artificial Intelligence 2 (1971) 189–208.

[7] Y. Gil, J. Blythe, PLANET: A sharable and reusable ontology for representing plans, in: Proceedings of the AAAI 2000 Workshop on Representational Issues for Real-World Planning Systems, AAAI Press, Menlo Park, CA, 2000, pp. 28–33.
[8] P. Hayes, The second naïve physics manifesto, in: J. Hobbs, R. Moore (Eds.), Formal Theories of the Commonsense World, Ablex, Norwood, NJ, 1985, pp. 1–36.
[9] D. McDermott, The 1998 AI planning systems competition, AI Magazine 21 (2) (2000) 35–55.
[10] A. Newell, H. Simon, GPS: A program that simulates human thought, in: E. Feigenbaum, J. Feldman (Eds.), Computers and Thought, McGraw-Hill, New York, 1963, pp. 279–293.
[11] C. Ortiz, A commonsense language for reasoning about causation and rational action, Artificial Intelligence 111 (2) (1999) 73–130.
[12] E. Pednault, ADL: Exploiting the middle ground between STRIPS and the Situation Calculus, in: R. Brachman, H. Levesque (Eds.), Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning, vol. 1, Morgan Kaufmann, San Francisco, CA, 1989, pp. 324–332.
[13] D. Wilkins, Practical Planning: Extending the Classical AI Planning Paradigm, Morgan Kaufmann, San Francisco, CA, 1988.