



# Probabilistic Selection of Case-Based Explanations in an Underwater Mine Clearance Domain

Venkatsampath Raja Gogineni<sup>(✉)</sup>, Sravya Kondrakunta,  
Danielle Brown, Matthew Molineaux, and Michael T. Cox

Wright State University, Dayton, OH 45435, USA  
gogineni.l4@wright.edu

**Abstract.** Autonomous agents should formulate and achieve goals with minimum support from humans. Although this might be feasible in a perfectly static world, it is not as easy in the real world where uncertainty is bound to occur. One approach to solving such a problem is to formulate goals based on cases that explain discrepancies observed in the environment. However, in an uncertain world, multiple such cases often apply (i.e., as alternative explanations). Moreover, agents in the real world often have limited resources to achieve their missions. So, it is risky to generate and achieve goals for every applicable explanatory case. Our solution to these problems is to down-select the retrieved cases based on probabilities derived using Bayesian inference, then to monitor the selected cases' validity based on observed evidence. We evaluate the performance of an agent in an underwater mine clearance domain and compare it to another agent that selects a random case from the candidate set.

**Keywords:** Case selection · Case-base explanation · Explanation patterns

## 1 Introduction

Agents in a mine clearance domain must perform critical surveillance tasks with high accuracy in waters where communication and observability are limited. Due to unpredictable and dangerous events such as explosions in unexplored areas, intelligent behavior is required to understand and respond to the environment. In this domain, explanation is useful for both monitoring the environment and engendering trust in human operators who have only intermittent contact with the agent. Trust is not investigated in this paper. However, we consider the problem of selecting an explanatory case from a candidate set of applicable cases for a deliberative mine hunting agent that must respond to the discrepancies.

Our agent for this domain is called *GATAR (Goal-driven Autonomy for Trusted Autonomous Reasoning)* [1]. GATAR plans to achieve its goals, then executes each step in this plan after checking to confirm that its preconditions are met. The actions and the postconditions constitute GATAR's *expectations* about the world. When they do not match the current observations of the world, GATAR tries to recognize the cause of the discrepancy and predict its effects on the agent's goals. However, when

there are multiple cases that might explain a discrepancy, GATAR intelligently selects a case based on the observations it possesses. Furthermore, it also forms *expectations* based on the selected explanatory case to monitor its validity. These cognitive capabilities provide three benefits: first, they help GATAR to intelligently respond to such events and prevent their recurrence; second, they help GATAR adapt its reasoning behind explanation selection to observed evidence; third, they help GATAR communicate the rationale behind its behavior to third parties. This third benefit is critical to building trust when working with humans [2]. While the GATAR agent is the primary focus of this paper, we expect lessons learned and results to be generalizable; we expect intelligent explanation-based behavior with deliberately selected goals to be useful in other critical domains like surveillance, medicine and autonomous driving.

A different approach to responding to discrepancies would be to generate contingent plans in advance that cover all possibilities. Unfortunately, this is computationally intractable in most domains and handling all contingencies for an unexpected event can be overwhelming. Our approach requires additional domain knowledge, but any specific domain of interest, an abstract case-base of explanations defined by domain experts can be retrieved and adapted to explain a discrepancy.

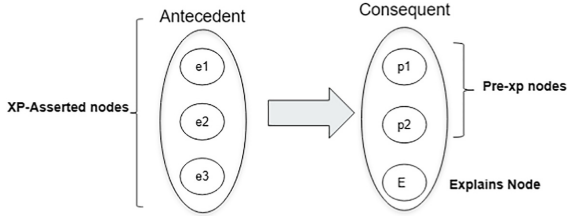
We present a Bayesian approach to select an explanation among the candidate explanatory cases retrieved from the case base. Moreover, expectations are extracted from the selected explanation to monitor its viability. Here explanation provides a causal basis for goal generation and enables creation of communicative rationales for goal changes to third parties. This approach follows *Goal Directed Autonomy (GDA)* principles [3–7], in which agents respond to *discrepancies* (i.e., agent expectation failures) by generating explanations and generating goals based on those explanations.

In Sect. 2, we describe the representation of the explanatory cases, their retrieval, selection, goal formulation and GATAR’s algorithmic approach to a discrepancy. A description of the domain, possible explanatory cases in the domain, and their retrieval are followed in Sect. 3. Section 4 presents the working example of the GATAR agent in a sample scenario. Evaluation and empirical results are presented in Sects. 5 and 6. Related work is illustrated in Sect. 7. Finally, the conclusion completes the paper in Sect. 8.

## 2 Case-Based Explanation Patterns

In our work, we use case-based explanations [8–10]. Each case in the case-base is an abstract *explanation pattern (XP)* [11, 12] engineered for a specific domain (see Fig. 1). An XP is a data structure that represents a causal relationship between two states and/or actions; each action/state is abstractly defined with variables to be adapted during or after case retrieval. In GATAR, an action or state is referred to as a *node* and different types of nodes are described based on their role in an XP.

- *Explains node*: A discrepancy/unknown state that is observed;
- *Pre-XP node*: Action/state that is observed along with the explains node;
- *XP-asserted node*: Action, state or XP contributing to the explanation’s cause.



**Fig. 1.** The explanation pattern (XP) causal structure in which XP-asserted nodes (e1, e2, e3) form an antecedent, and a consequent is made up of pre-XP nodes (p1, p2, p3) and an explains node (E); XP-asserted nodes thus cause the associated explains and pre-XP nodes.

### 2.1 GATAR’s Algorithm to Respond to the Discrepancies

Algorithm 1 represents GATAR’s approach toward identifying and responding to a discrepancy. Given a plan  $\pi = \langle a_1 \dots a_n \rangle$  to achieve goals from agenda  $\hat{G} = \{g_1 \dots g_m\}$ , whenever its current observations ( $s_c$ ) do not match the expectations ( $s_e$ ), a discrepancy is detected. These observations ( $s_c$ ) are obtained (line 1) from the successor function ( $\gamma$ ) that takes in the current state ( $s_c$ ) and action ( $a_1$ ) [13]. Since, GATAR is currently executing  $a_1$ , its plan is updated to the set of remaining actions (line 2). GATAR then adds to its expectations ( $s_e$ ) the preconditions ( $a_1^+$ ) and effects ( $a_1^-$ ) of the current action (line 3). When expectations differ from observations (line 4), the algorithm tries to explain this discrepancy from the case base (lines 5–8). First, a candidate set of explanations  $c = \{\chi_1 \dots \chi_k\}$  is retrieved from the case-base  $c = \{\chi_1 \dots \chi_k \dots \chi_l\}$  of explanations. An explanatory-case ( $\chi_s$ ) is selected from these candidates by applying Bayesian inference (line 5). Next, additional expectations are extracted from the XP-asserted nodes of the selected explanation and added to the current set of GATAR’s expectations (line 6). This facilitates monitoring the validity of the explanatory case. The interpretation function (beta) sets a new current goal set to respond to the discrepancy (line 7) (see [32]). Finally, the new goals ( $g_c$ ) are added to the goal agenda ( $\hat{G}$ ) (line 8).

**Algorithm 1:**

*DiscrepancyCheck* ( $\pi, s_c, s_e, \hat{G}, c, g_c$ )

1.  $s_c \leftarrow \gamma(s_c, a_1)$
2.  $\pi \leftarrow \langle a_2 \dots a_n \rangle$
3.  $s_e \leftarrow s_e \cup pre(a_1) \cup a_1^+ - a_1^-$
4. if  $s_c \neq s_e$
5.  $\chi_s \leftarrow Select(Retrieve(c, s_c))$
6.  $s_e \leftarrow s_e \cup XP\text{-asserted}(\chi_s)$
7.  $g_c \leftarrow \beta(s_c, g_c)$
8.  $\hat{G} \leftarrow \hat{G} \cup g_c$

## 2.2 Retrieving, Reusing and Revising Explanation Patterns from a Case Base

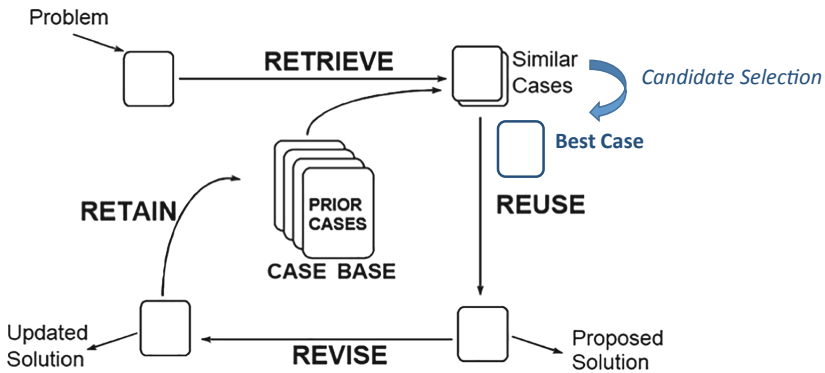


Fig. 2. The CBR process with candidate selection (adapted from [14])

Case-based reasoning follows a four-step process to retrieve, reuse, revise and retain cases [14–16] (see Fig. 2). In the current work, we assume that all cases are defined by domain experts, so we do not consider retention. The following describes how XPs are retrieved, reused and revised.

A set of abstract XPs is retrieved when an unpredicted state or action observed by the agent unifies with each explains node of an XP in the case-base. If the unification turns out to be successful then the pre-XP nodes of the corresponding case are unified with the observations of the corresponding states or actions, if they turn out to be successful then the specific XP is retrieved. The retrieved abstract XP is reused by binding variables in the antecedent to values found during unification of the consequent. However, if the XP-asserted nodes in the reused XP contain hypothetical information they can be revised when the new knowledge is obtained from further observations. We now describe our approach to selecting an explanation from a retrieved candidate set.

## 2.3 Probabilistic Selection of Explanation Patterns from a Candidate Set

In our work, we use Bayesian inference to select an explanation from the candidate set of retrieved explanations. Bayesian inference takes prior knowledge about the parameter and uses newly collected data or information to update the prior beliefs [17]. The agent uses its observations of states/actions as new information in updating its prior beliefs. Bayes equation adapted to the explanation patterns is given as follows:

$$P(XP|evidence) = \frac{P(evidence|XP) * P(XP)}{P(evidence)}$$

$$P(\text{evidence}|\text{XP}) = \frac{\text{no. of times the evidence obtained given explanation}}{\text{total no. of times evidences obtained given explanation}}$$

$$P(\text{XP}) = \frac{\text{no. of times the explanation is selected}}{\text{total no. of times the explanation is in a candidate set}}$$

$$P(\text{evidence}) = \frac{\text{no. of times the evidence is obtained}}{\text{total no. of times evidence is in a candidate set}}$$

Each XP-asserted node is considered as an evidence, so Bayes equation is applied for every XP-asserted node. The probability of the explanation given all evidence is

$$P(\text{XP}|\text{evidence}_{1..n}) = P(\text{XP} | (\text{evidence}_1 \cap \text{evidence}_2 \dots \text{evidence}_n))$$

For all the explanations, the one with the highest  $P(\text{XP}|\text{evidence}_{1..n})$  is the one to be selected. However, since our explanations are designed manually by domain experts, we assume the domain experts provide the expected values for different prior probabilities i.e.  $P(\text{XP})$ ,  $P(\text{evidence}_{1..n})$ , before the mission starts. Moreover, if there is no evidence obtained then an explanation having the highest  $P(\text{XP})$  among the candidate cases is selected. Finally, goals are formulated and the  $\text{XP}$  monitored to check validity.

## 2.4 Goal Formulation and Monitoring Explanation Patterns

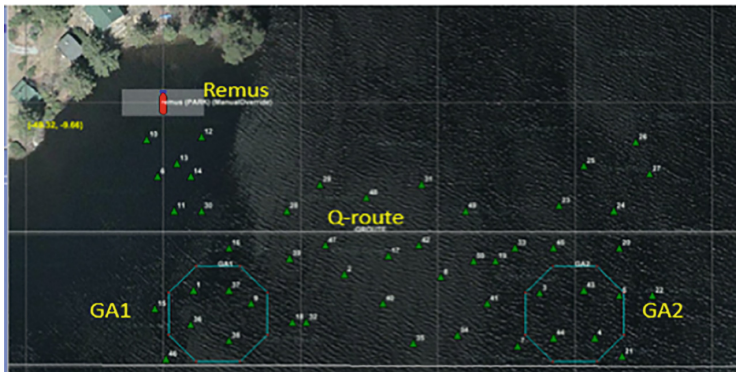
Goal formulation is essential for an intelligent agent to respond to discrepancies [18, 19]; in GATAR, we perform formulation by preventing the recurrence of one or more explanation antecedent nodes. Antecedent nodes may include actions and/or states; therefore, when GATAR wishes to prevent an undesired consequent from recurring, it considers the elimination of antecedent actors or objects that participate in antecedent states as potential goals. The potential goals are generated using the removal mapping function that takes in the actors or objects and outputs the goals that eliminate them.

Monitoring a selected explanation is essential for an intelligent agent to adapt its beliefs and misclassifications. In our work, each node in the XP-asserted nodes of the selected  $\text{XP}$  is added to the agent's expectations. These nodes constitute the evidence. Whenever the evidence matches the agent's observations of the world, the agent changes its beliefs. However, when the evidence contradicts the agent's observations, the selected  $\text{XP}$  is switched with the next probable explanation.

## 3 Underwater Mine Clearance Domain

Our approach is implemented in a limited mine clearance domain [20], which is simulated using MOOS-IVP [21], software that provides complete autonomy for marine vehicles. Figure 3 shows the simulation of the mine clearance domain with a GATAR agent directing a Remus unmanned underwater vehicle. The Q-route is a safe

passage for ships to enter and leave the port and is represented as a rectangular area in Fig. 3. GA1 and GA2 are the two octagonal areas where mines are expected to exist, while the triangular objects are the mines. The goals of the agent are to survey and clear mines in GA1 and GA2. These goals are given to the agent after a reconnaissance mission performed by a different agent across the whole sea route. The Remus has a sonar sensor with a specific width of ten units and a length of five units to detect mines.



**Fig. 3.** Underwater Mine Clearance domain with two clearance areas in the Q-route. (labelled items in this figure match later figures)

### 3.1 Discrepancies in Underwater Mine Clearance Domain

In the underwater mine clearance domain, several events often co-occur simultaneously, and many events cannot be predicted based on knowledge available to an agent. These events might affect the agent itself or the mission of the agent. Explanations help the agent to recognize these events and respond to them. We will look at several uncertain events that might happen.

Events in this domain include minelaying, sensor failure, and reconnaissance failure. Minelaying events occur when an enemy ship, aerial vehicle, or fishing vessel lays traps to hurt friendly ships. Removing such mines within areas GA1 and GA2 is an explicit goal for GATAR in the above scenario. Sensor failure event indicates that a faulty GATAR's sensor is responsible for a misclassification of mine, and the failure of proper reconnaissance mission indicates that an agent prior to GATAR did not identify mines which in turn failed its mission. We will look at some of the explanations in the next section that sheds some knowledge on the discrepancies that happen in this domain.

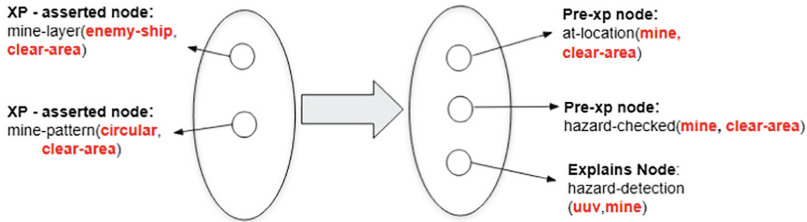
### 3.2 Plausible Explanations in Underwater Mine Clearance Domain

In GATAR, explanations are retrieved by a version of the *Meta-AQUA* component [22], a story understanding system that tries to explain discrepancies in a story through

use of case-based explanations. We have integrated this system with the *MIDCA* (*Metacognitive Integrated Dual Cycle Architecture*) component [23], a cognitive architecture that perceives and acts directly on the world, to examine the interaction between explanation generation (by Meta-AQUA) and goal formulation (provided by MIDCA); for the purposes of this paper, we refer to the combined system as the GATAR agent.

As mentioned earlier, each explanation in our case-base is abstractly designed by the domain experts to cover the possible discrepancies in the underwater mine clearance domain. We will look at one of the detailed candidate explanatory case that can occur for a discrepancy of detecting multiple mines at a location.

Figure 4 represents an abstract XP structure, which explains that an enemy ship laid the mines in the clear-area and hence the UUV (unmanned underwater vehicle) detected multiple mines. UUV is an abstraction for GATAR, while Clear-area is an abstraction for areas that are not expected to have mines. As described earlier the XP structure is in the form of an antecedent causing consequent. The consequent contains the explains node “hazard-detection(uuv, mine)” to represent the discrepancy of detecting a mine by the UUV, while the Pre-XP nodes “at-location(mine, clear-area)” and “hazard-checked(mine, clear-area)” are the observational support to the discrepancy. These Pre-XP nodes thus convey that a mine is already checked prior to the currently detected mine, and implicitly conveys that multiple mines exist.



**Fig. 4.** XP Structure that describes enemy ship caused multiple mines

The antecedent contains the XP-asserted nodes of mine laying activity by the enemy-ship in the clear-area represented by the action “mine-layer(enemy-ship, clear-area)” and the mine pattern being circular is represented by state “mine-pattern(circular, clear-area)”. Whenever this XP is retrieved the abstracted variables will be replaced by the observations. For example, UUV will be replaced by GATAR, clear-area will be replaced by all the areas in the domain except GA1 and GA2.

In a similar structure there are multiple explanations for multiple discrepancies that might happen in the domain. Below are the high-level descriptions of those explanations along with their responsive behaviors in Table 1.

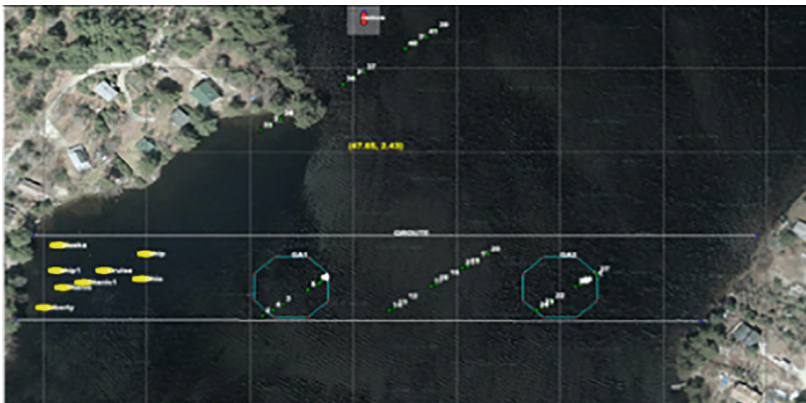
**Table 1.** Explanations for the discrepancies along with the behaviors

Explanation	Discrepancy	Behavior
<i>Fisher-XP</i> : Fisher-vessel laid a single mine	Single mine detected at the clear-area	Remove the single mine (Q-route) and report existence of fishing vessel
<i>Sensor-XP</i> : GATAR’s sensor failure	Single mine detected at the clear-area	Recalibrate the sensor and continue the mission
<i>Tide-XP</i> : Mines drifted from expected regions GA1 or GA2 with tides	Single or multiple mines detected at clear-area	Clear the mines and continue the mission
<i>Reconnaissance-XP</i> : Reconnaissance failed	Multiple mines detected in the clear area	Survey and clear the mines in the whole Q-route
<i>Enemy-Ship-XP</i> : Enemy ship laid the circular pattern of mines	Multiple mines detected in the clear area	Survey region covering circular area with calculated radius & clear mines
<i>Enemy-Aerial-XP</i> : Aerial vehicle laid the line pattern of mines	Multiple mines detected in the clear area	Survey region across line region with statistically calculated slope & clear mines

### 4 Example of Selecting a Case-Based Explanation

An example scenario from the underwater mine clearance domain can help us understand the application of selecting explanations from the candidate set by GATAR. Moreover, this will also give an idea of how such an application can improve the performance of GATAR.

Figure 5 represents the example scenario in the underwater mine clearance domain, where there are four minefields: first, at GATAR’s transit to the Q-route; second, at the GA1; third, at GATAR’s area of transit from GA1 to GA2; and fourth, at the GA2. However, GATAR has a mission to clear the second and fourth minefields, so any mines encountered in the first and the third minefields are discrepancies. When GATAR encounters the first mine in the first minefield, it retrieves a candidate set of explanatory cases. Table 2 shows this set of explanatory cases retrieved as well as the respective selection probabilities.



**Fig. 5.** A scenario where mines are laid by an aerial vehicle



These probability values were chosen to match typical values in the domain. Since Fisher-XP is the only explanation with evidence, the probability to select an explanation is 0.5 and the probabilities of the Sensor-XP and Tide-XP are 0.14 and 0.08 respectively. Thus Fisher-XP is selected due to its higher probability and the table is updated.

Table 3 shows the updated probabilities calculated using the Bayesian inference as described in the Sect. 2.3. Furthermore, a goal is generated to report the existence of a fisher vessel laying mines after the mission. Finally, the evidence is added to GATAR's expectations i.e. there is only one mine in the transit area to the Q-route.

**Table 2.** Probabilities of the retrieved explanations

Explanation ( <i>XP</i> )	P ( <i>XP</i> )	P ( <i>XP</i>   evidence)
<i>Fisher-XP</i> : Fisher vessel laid a single mine	0.20	P ( <i>XP</i>   single-mine) = 0.5
<i>Sensor-XP</i> : GATAR's sensor failure	0.14	No evidence obtained regarding damage of sensor
<i>Tide-XP</i> : Mines drifted by the tides from the mine expected regions (GA1 or GA2)	0.08	No evidence obtained regarding tides and the mines not at expected minefield locations

**Table 3.** Probabilities after selecting the explanation

Explanation ( <i>XP</i> )	P ( <i>XP</i> )	P ( <i>XP</i>   evidence)
<i>Fisher-XP</i> : Fisher vessel laid a single mine	0.207	P ( <i>XP</i>   single-mine) = 0.509
<i>Sensor-XP</i> : GATAR's sensor failure	0.138	No evidence obtained regarding damage of sensor
<i>Tide-XP</i> : Mines drifted by the tides from the mine expected regions (GA1 or GA2)	0.079	No evidence obtained regarding tides and the mines not at expected minefield locations

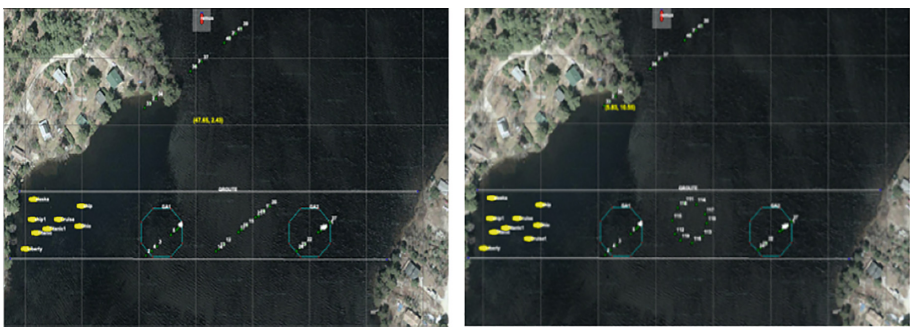
When GATAR encounters another mine in the same area, its expectation of a single mine is violated, and it retrieves another set of candidate explanatory cases to explain the discrepancy. Since the previously selected explanation is not valid, it updates its probabilities and drops the goal to report about the fisher vessel. Similarly, as described above, GATAR selects the explanation that an aerial vehicle laid the mines, updates its probabilities, formulates goals to report about the vehicle and finally generates an expectation that the mines exist in a straight line.

After clearing all the mines from GA1, GATAR encounters mines in the third minefield, selects the explanation that the mines exist in a straight line, clears all the mines in the Q-route, and updates its probabilities. Finally, it returns to the GA2 and clears all the mines in the GA2. Later after the mission GATAR reports the existence of the aerial vehicle and its behavior to the base which is outside the scope of the paper. However, in this scenario, GATAR intelligently made the Q-route safe for the ships to traverse.

## 5 Experimental Setup

As previously mentioned, GATAR retrieves explanatory cases that explains any discrepancies. A case-base of ten abstract explanatory cases are used to cover all the behaviors of the GATAR agent in this domain. From the set of the retrieved cases it selects an applicable explanation by applying Bayesian inference. Furthermore, the antecedents of the candidate explanations are monitored to update the beliefs of the agent. GATAR's ability to the above intelligent behavior is evaluated by number of ships that traverse the Q-route without hitting mines. Moreover, the performance of GATAR is compared to a random agent. The random agent is like the GATAR agent in retrieving a candidate set of explanatory cases. However, it differs from GATAR in the selection process. It selects a random explanation from the candidate set to respond to a discrepancy. The experiment is laid out in terms of two scenarios that are differentiated by the placement of mines. Each scenario has three groups of three ships that start at a specific location and ends across the other side of the Q-route at another specific location. The first group will start with incremental delays while the second and third groups will start with a constant delay of 0.25 and 0.50 min respectively following the first group. The agent has the goals to clear mines in the areas GA1 and GA2.

Figure 6 represents the experimental setup of both scenarios 1 and 2. In the first scenario, an aerial vehicle laid mines in a line pattern in the areas of transit, GA1, GA2 and the transit area between GA1 and GA2. The setup gives the agent a proper evidence that an aerial vehicle laid the mines while pursuing its goals. In the second scenario, an enemy ship laid the mines in the area between GA1 and GA2, while the aerial vehicle laid the mines at transit, GA1 and GA2. The first scenario tests GATAR's ability to select the correct explanation by obtaining evidence while the second scenario tests GATAR's ability to switch explanations to clear mines between GA1 and GA2 when the evidence implies an explanation that an aerial vehicle is laying mines.



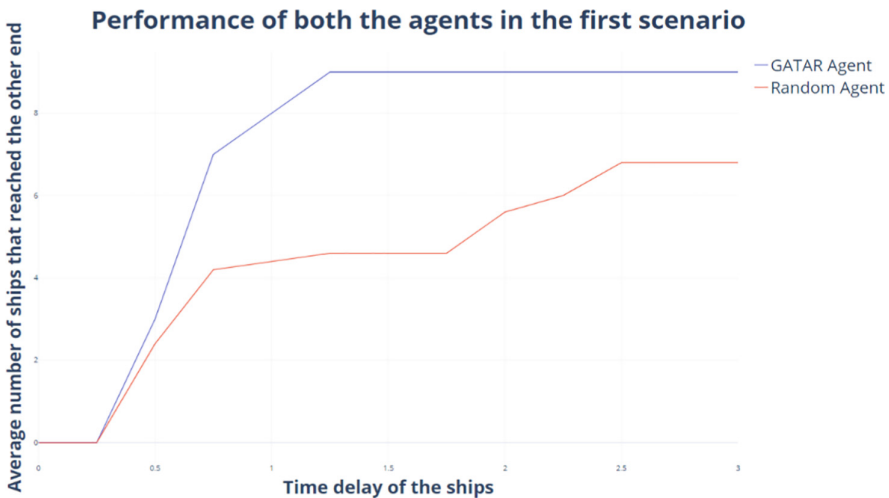
**Fig. 6.** (Left) Scenario 1 – mines laid by an aerial vehicle. (Right) Scenario 2 – mines laid by both an aerial vehicle and an enemy ship.

## 6 Empirical Results

Figure 7 shows the comparison of the performance with GATAR and the random agent in the scenario 1. The X-axis shows the starting time, in increments of 0.25 min, for the first group of ships after the agent's mission starts. The Y-axis is the average number of ships that successfully traversed through the Q-route. The experiment runs five times for every time delay and for every agent and then averaged for each experiment. GATAR outperforms the random agent at every time interval greater than 0.25.

When we look closely into results with a delay of 0.25 min, when the first group of ships start, both the agents are still in their transit to the GA1. By the time the third batch of ships start, both the agents cleared mines in GA1 and are in transit to Q-route. This implies that a delay of 0.25 min is not long enough for both agents to clear enough mines in the Q-route for any ship to survive. However, at an interval of 0.5 min of time, both the agents could clear mines in GA1 as well as some on their transit to GA2 before the third group of ships started, which signifies the steep increase in the curve. The random agent underperforms because of the failure to select the correct explanation leading to a different behavior.

At 0.75 interval of time, the GATAR agent could successfully clear GA1 and some mines on its transit to the GA2 by the time the second group of ships start. The GATAR agent completely clears all the mines in the Q-route before the third batch of ships start. The random agent underperformance is due to its wrong choice in behavior. Finally, at 1.25 interval of time, GATAR could completely clear all the mines in the Q-route, so all the ships survived. Even at 2.5 intervals of time, the random agent could not clear all the mines because of its selection of behaviors that ignore the mines on their transit to the Q-route.



**Fig. 7.** Performance of the GATAR and the random agent in scenario 1

Figure 8 compares the performance of GATAR and the random agent in scenario 2. The experiment pattern follows the same as described above. GATAR outperforms the random agent at every interval of time greater than the 0.75. In this scenario, GATAR clears the mines in GA1 and when it happens to identify a mine between GA1 and GA2 in Q-route, it generates a behavior to traverse in a line based on the prior evidence it obtained. After observing that the pattern is not a line, the GATAR agent changes its behavior to a deep search pattern. Later it continues to clear the mines in GA2.

At 0.25 min of time, the random agent happens to select a behavior to clear only the mines that it came across during two of the five random runs before the third group of ships started. This resulted in clearing a mine at the upper part of the circle of mines between the areas GA1 and GA2 in the Q-route, which allowed one of the ships to survive. However, this is not the case with GATAR where all ships sank at time .25. At 1.25 min, GATAR clears all mines in the Q-route which allowed all ships to traverse the route. At a delay of 2.5 min of time, the random agent seems to complete all its goals before the ships start. However, because of its wrong choice of explanations, the formulated goals could not allow for safe passage for all the ships.

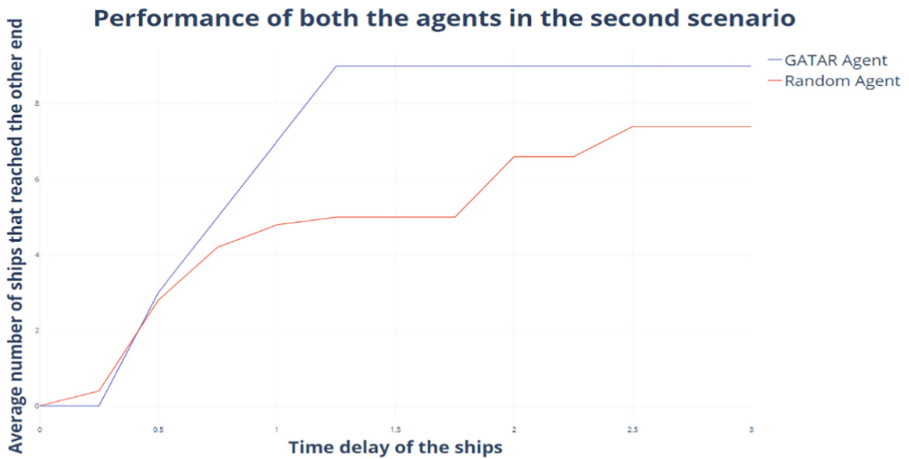


Fig. 8. Performance of the GATAR and the random agent in scenario 2

## 7 Related Work

Explanation patterns were introduced by Schank in 1982 [8] and were later used in the story understanding systems SWALE [24, 25] and AQUA [26]. SWALE is a case-based approach to explanation of discrepancies in a story that retrieves, adapts and stores explanation patterns. SWALE demonstrated an early technique for ruling out competing explanations using memory knowledge. AQUA (Asking Questions and Understanding Answers) operates by first questioning missing knowledge in a story, then using explanation patterns to understand the answers.

Gentner and Forbus in [27], present the MAC/FAC approach, which closely aligns to our approach. It is a two-phase retrieval process to improve the performance of the retrieval process. MAC (Many are chosen) refers to using flat structures to eliminate irrelevant cases outright while FAC (Few are chosen) refers to applying computationally complex algorithms to rank the cases from MAC. Our approach aligns closely with the two-phase retrieval process, where we select a candidate set of *XP*'s using explains and Pre-*XP* nodes and apply probabilities to select a single *XP*. Moreover, Morwick and Leake in [28], show a performance increase in having such two-phase retrieval approach.

Roth-Berghofer et al's [29] work on classifying explanations and their use-cases according to the user's intentions is one of the theoretical research directions towards explanations in case-based reasoning (see also [30]). This paper introduces the concept of "explanation goals" that are used to decide when and what the system should explain to users based on their expectations. In future research, we will investigate application of these techniques to prevent the system from repeatedly explaining the same type of unexpected events to a user who is already familiar with them. This paper also talks about different kinds of explanations and classifies them into four different knowledge containers, all of which are used to generate explanations based on the user's goals or intentions.

In [2], a robot adapts its behavior to gain trust in human machine teaming using the approach of case-based reasoning. In addition, Floyd and Aha [31] presented an approach to explain such adaptations based on an operator's feedback, and evaluated their system based on how closely the explanations aligned with the operator's feedback. Our interest in generating explanations of the intelligent behavior of an agent aligns closely with the interests of this paper.

## 8 Conclusion

In this paper we discussed a probabilistic approach of selecting an explanatory case from the candidate set of retrieved cases for the discrepancy. Moreover, we have also presented an approach to monitor the selected case, which helps GATAR to adapt its beliefs and switch cases if a selection error occurs. Finally, the results show that the performance of GATAR is better than the random agent. The causal structure of explanations helps GATAR communicate and justify its behavior to human users.

In some cases, there can be more than one explanation relevant to a discrepancy. For example, if both an enemy ship and an aerial vehicle laid mines in the same area then it would be incorrect to choose one explanation. In future research, we would like to reason about the probability of co-occurrence of causal events leading to discrepancies.

We also acknowledge that our current explanatory cases will provide only abstract mine patterns which are evenly spaced or uniformly distributed. However, the real world contains noise leading to misclassification of the evidence. In the future, we would like to use statistical learning algorithms to predict mine patterns as well as their distributions. Furthermore, GATAR should reason about the tradeoff between the time required to clear the mines in GA1 and GA2 and the time to pursue its additional goals

from the selected explanatory case. We expect that such reasoning will improve GATAR's performance.

Moreover, we also want GATAR to explain the rationale behind its intelligent behavior to the human operator and obtain some feedback related to the hypothetical evidence. This will improve the quality of explanatory cases selected from the case-base.

Finally, we want GATAR to reason about the tradeoff between immediately formulating goals from a selected explanatory case and formulating goals after obtaining evidence. We expect this functionality to help GATAR to adapt after selection failures.

**Acknowledgements.** This research was supported by AFOSR under grant FA2386-17-1-4063 and by ONR under grant number N00014-18-1-2009. We thank the anonymous reviews for the comments and suggestions.

## References

1. Gogineni, V., Kondrakunta, S., Molineaux, M., Cox, M.T.: Application of case-based explanations to formulate goals in an unpredictable mine clearance domain. In: Proceedings of the ICCBR-2018 Workshop on Case-Based Reasoning for the Explanation of Intelligent Systems, Stockholm, Sweden, pp. 42–51 (2018)
2. Floyd, M.W., Drinkwater, M., Aha, D.W.: Trust-Guided Behavior Adaptation Using Case-Based Reasoning. Naval Research Laboratory, Washington, United States (2015)
3. Paisner, M., Cox, M., Maynard, M., Perlis, D.: Goal-driven autonomy for cognitive systems. In: Proceedings of the Annual Meeting of the Cognitive Science Society, vol. 36, no. 36 (2014)
4. Molineaux, M., Klenk, M., Aha, D.W.: Goal-driven autonomy in a navy strategy simulation. In: AAI, pp. 1548–1554 (2010)
5. Dannenhauer, D., Munoz, A.H.: Raising expectations in GDA agents acting in dynamic environments. In: IJCAI, pp. 2241–2247 (2015)
6. Cox, M.T.: Goal-driven autonomy and question-based problem recognition. In: Proceedings of the 2nd Annual Conference on Advances in Cognitive Systems, Maryland, USA, pp. 29–45 (2013)
7. Munoz-Avila, H., Aha, D.W., Jaidee, U., Klenk, M., Molineaux, M.: Applying goal driven autonomy to a team shooter game. In: FLAIRS Conference (2010)
8. Schank, R.C., Kass, A., Riesbeck, C.K.: Inside Case-Based Explanation. Psychology Press, London (2014)
9. Cox, M.T., Burstein, M.H.: Case-based explanations and the integrated learning of demonstrations. *Künstliche Intelligenz (Artif. Intell.)* **22**(2), 35–38 (2008)
10. Ram, A.: Indexing, elaboration and refinement: incremental learning of explanatory cases. *Mach. Learn.* **10**, 201–248 (1993)
11. Schank, R.C.: Explanation Patterns: Understanding Mechanically and Creatively. Psychology Press, London (2013)
12. Ram, A.: A theory of questions and question asking. *J. Learn. Sci.* **1**(3 and 4), 273–318 (1991)
13. Ghallab, M., Nau, D., Traverso, P.: Automated Planning: Theory and Practice. Elsevier, Amsterdam (2004)
14. de Mántaras, R.L., et al.: Re-trieval, reuse and retention in case-based reasoning. *Knowl. Eng. Rev.* **20**(3), 215–240 (2006)

15. Aamodt, A., Plaza, E.: Case-based reasoning: foundational issues, methodological variations, and system approaches. *AI Commun.* **7**(1), 39–52 (1994)
16. Kolodner, J.: *Case-Based Reasoning*. Morgan Kaufmann, San Francisco (1993)
17. Box, G.E., Tiao, G.C.: *Bayesian Inference in Statistical Analysis*. Wiley, Hoboken (2011)
18. Maynard, M., Cox, M.T., Paisner, M., Perlis, D.: Data-driven goal generation for integrated cognitive systems. In: 2013 AAAI Fall Symposium Series (2013)
19. Hanheide, M., et al.: A framework for goal generation and management. In: *Proceedings of the AAAI Workshop on Goal-Directed Autonomy* (2010)
20. Kondrakunta, S., Gogineni, V., Molineaux, M., Munoz-Avila, H., Oxenham, M., Cox, M.T.: Toward problem recognition, explanation and goal formulation. In: *Proceedings of the 6th Goal Reasoning Workshop at IJCAI/FAIM-2018, Stockholm, Sweden* (2018)
21. Benjamin, M.R., Schmidt, H., Newman, P.M., Leonard, J.J.: Nested autonomy for unmanned marine vehicles with MOOS-IvP. *J. Field Robot.* **27**(6), 834–875 (2010)
22. Cox, M.T., Ram, A.: Introspective multistrategy learning: on the construction of learning strategies. *Artif. Intell.* **112**(1–2), 1–55 (1999)
23. Cox, M.T., Alavi, Z., Dannenhauer, D., Eyorokon, V., Munoz-Avila, H., Perlis, D.: MIDCA: a metacognitive, integrated dual-cycle architecture for self-regulated autonomy. In: *AAAI* (2016)
24. Schank, R.C., Leake, D.B.: Creativity and learning in a case-based explainer. *Artif. Intell.* **40**(1–3), 353–385 (1989)
25. Leake, D.B.: *Evaluating Explanations: A Content Theory*. Psychology Press, London (2014)
26. Ram, A.: *AQUA: Questions that drive the explanation process*. Georgia Institute of Technology (1993)
27. Gentner, D., Forbus, K.: MAC/FAC: A model of similarity-based retrieval. In: *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society, Chicago, IL*, pp. 504–550 (1991)
28. Kendall-Morwick, J., Leake, D.: A study of two-phase retrieval for process-oriented case-based reasoning. In: Montani, S., Jain, L.C. (eds.) *Successful Case-Based Reasoning Applications-2*, vol. 494, pp. 7–27. Springer, Heidelberg (2014). [https://doi.org/10.1007/978-3-642-38736-4\\_2](https://doi.org/10.1007/978-3-642-38736-4_2)
29. Roth-Berghofer, T.R., Cassens, J.: Mapping goals and kinds of explanations to the knowledge containers of case-based reasoning systems. In: Muñoz-Ávila, H., Ricci, F. (eds.) *ICCBR 2005. LNCS (LNAI)*, vol. 3620, pp. 451–464. Springer, Heidelberg (2005). [https://doi.org/10.1007/11536406\\_35](https://doi.org/10.1007/11536406_35)
30. Aamodt, A.: Explanation-driven case-based reasoning. In: Wess, S., Althoff, K., Richter, M. (eds.) *Topics in Case-Based Reasoning*, vol. 837, pp. 274–288. Springer, Berlin (1994). [https://doi.org/10.1007/3-540-58330-0\\_93](https://doi.org/10.1007/3-540-58330-0_93)
31. Floyd, M.W., Aha, D.W.: Incorporating transparency during trust-guided behavior adaptation. In: Goel, A., Díaz-Agudo, M., Belén, Roth-Berghofer, T. (eds.) *ICCBR 2016. LNCS (LNAI)*, vol. 9969, pp. 124–138. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-47096-2\\_9](https://doi.org/10.1007/978-3-319-47096-2_9)
32. Cox, M.T.: A model of planning, action, and interpretation with goal reasoning. *Adv. Cogn. Syst.* **5**, 57–76 (2017)