Advances in Cognitive Systems 3 (2014) 123–142 Submitted 9/2013; published 7/2014

Anomaly-Driven Belief Revision and Noise Detection by Abductive Metareasoning

Joshua Eckroth ECKROTH@CSE.OHIO-STATE.EDU
John R. Josephson JJ@CSE.OHIO-STATE.EDU
Department of Computer Science and Engineering, The Ohio State University, Columbus, Ohio 43210 USA

Abstract

Cognition may reasonably be distinguished into world-estimation and planning tasks. Our focus in this work is the world estimation task, i.e., the task of establishing and updating beliefs about the world. One aspect of being intelligent in this task is noticing one's mistakes and correcting them. An intelligent system may be realized by dividing its operation into a base-level reasoning system and a metareasoning system. The base-level system is responsible for processing inputs from the world and recording its conclusions in a belief state. The metareasoning system monitors the base-level system so that it can detect symptoms of errors in the belief state and attempt belief revisions. We describe and evaluate such a system in this report. The base-level system is an abductive reasoner responsible for finding explanations for inputs given as reports of putative observations. When no plausible, consistent explanation is forthcoming for some reports, we say these unexplainable reports are anomalous. The presence of anomalies is a symptom of errors in the belief state since, in the usual case, all reports should be explainable. However, sometimes the anomalous reports are not reports of observations but rather false or noisy reports. An abductive metareasoning system attempts to explain anomalies as errors of various kinds and makes the appropriate revisions. If a sufficiently plausible explanation is not found, an anomalous report is attributed to noise. We evaluate this two-level system in a pair of object tracking tasks, one simulated and one based on aerial surveillance. Both tasks are challenging due to limited sensor capabilities and a very high level of noise. The proposed two-level system realizes significantly more accurate belief states and better noise detection than a system that lacks abductive metareasoning.

1. Introduction

Cognitive systems that include a metareasoning or self-reflective component have enjoyed renewed interest in recent years, as evidenced by workshops such as the AAAI-2008 Metareasoning Workshop and the volume that followed, *Metareasoning: Thinking about thinking* (Cox & Raja, 2011). Such architectures typically consist of a base-level reasoner, which reasons about object-level concepts, and a meta-level reasoner that monitors and manipulates the base-level reasoner. In this work, we describe and evaluate a cognitive system that is divided into a *abductive* base-level reasoner and an *abductive* metareasoner. The base-level abductive reasoner receives reports of putative observations and attempts to infer their true explanations, i.e., what caused the phenomenon that was observed and reported. The abductive reasoning is domain-general reasoning, concerning itself with generic evidence and hypotheses rather than object-level phenomena. In this work, domain-specific concepts are encapsulated in a separate module that appears as a black box to the reasoners. Sometimes, no plausible, consistent explanations for some reports are generated by the domain-specific module, leaving those reports unexplainable. We call such reports *anomalous* and describe how they may be symptoms of errors in the system's world estimate.

The presence of anomalies activates the metareasoning system, which treats anomalies as evidence of possible errors and attempts to determine their causes. Thus, the metareasoning system is itself abductive, and uses the same machinery (algorithms and code) as the base-level reasoner. We enumerate the possible causes of anomalies that are specific to abductive base-level reasoners. Besides errors in the belief state, an anomaly might be caused by noise, i.e. a false report that is anomalous partly because it is false. Thus, an anomaly does not always imply an error in the belief state. In our system, abductive metareasoning concludes that an anomaly is the result of noise as a fallback explanation when no other explanation of the anomaly is found that is consistent or sufficiently plausible. Thus, noise is not explicitly detected by domain-specific report properties, but rather implicitly detected by way of domain-general contextual considerations. We demonstrate that this two-level abductive system is effective at detecting and correcting errors, and detecting noise, even when faced with limited sensor capabilities and a very high level of noise.

Although other systems include an anomaly-driven metareasoning component to guide belief or theory revision and learning (Schmill et al., 2011; Bridewell, 2004; Cox & Ram, 1999), the system discussed here is, to our knowledge, the first to apply abduction where both the base-level and metareasoning components use the same machinery. Some benefits of this combined system are its generality, simplicity, and effectiveness at its task.

The remainder of this paper provides details. We introduce the pattern of abductive reasoning in the next section. This is followed in Section 3 by a formal definition of *belief states* and various other features of our abductive reasoning system. Section 4 details an efficient abductive reasoning algorithm. Next, we explore abductive metareasoning in Section 5. This section outlines the possible causes of anomalies in the base-level abductive reasoning system, and their corresponding belief revisions. Section 6 describes the experimental domains and results of our evaluation. The final two sections discuss related work and plans for future research.

2. Abductive Reasoning

By *abduction*, and *abductive inference*, we mean reasoning that follows a pattern approximately as follows (Josephson & Josephson, 1994):

> D is a collection of data (findings, observations, givens).
> Hypothesis H can explain D (would, if true, explain D).
> No other hypothesis can explain D as well as H does.
> –
> Therefore, H is probably correct.

In a process of trying to explain some evidence, the object is to arrive at an explanation that can be confidently accepted. An explanation that can be confidently accepted is an explanation that can be justified as being *the best explanation* in consideration of various factors such as plausibility, consistency, and completeness, and in contrast with alternative explanations. Thus, an explanation-seeking process – an abductive reasoning process – aims to arrive at a conclusion that has strong abductive justification. We hope that readers recognize abductive reasoning is a distinct and familiar pattern, and has an intuitively recognizable evidential force. It is reasonable to say it is part of commonsense logic. It can be recognized in a wide range of cognitive processes including diagnosis, scientific theory formation, language comprehension, and perception.

For the purpose of grounding our discussion of abductive reasoning, consider an object tracking task. Suppose that cameras and low-level blob detection generate reports of the form, "at time t and location x, y, a blob of size w, h was detected." Reports make up the collection of observations, D. Suppose the goal of the system is to establish correct beliefs of the form, "from time t to $t+1$, object o moved from x_t, y_t to x_{t+1}, y_{t+1}." Hypotheses have the same structure as beliefs; they are, in essence, candidate beliefs. Each hypothesis h would, if true, explain two reports: a report at time t and a report at time $t+1$. A hypothesis posits that both reports are observations of the same object. Each hypothesis may have some kind of *plausibility*, or measure of goodness based on knowledge of average speed of objects, similarity of the two observations at time t and $t+1$, etc. Two hypotheses h_i and h_j may be incompatible or inconsistent with each other if they both claim that the same object moved to two different locations at the same time, assuming objects are not allowed to split. Abductive reasoning is the process of finding the best explanation $\{h_1, h_2, \ldots, h_n\}$ such that most or all reports are explained.

3. Belief States

In order to keep track of evidence, hypotheses, and the plausibility and status of each hypothesis, our system constructs a *belief state* as characterized in Definition 3.1. Note that, for simplicity's sake, we treat reports as hypotheses that explain nothing but are themselves considered unexplained if they are accepted. Such hypotheses are initially accepted, but may be rejected (ignored) during metareasoning if they are subsequently deemed to be noise.

Definition 3.1. A *belief state* is a tuple $B = (H, X, P, S, V, I)$, where $H = \{h_1, \ldots, h_n\}$ is a (finite) set of hypotheses and each hypothesis is a grounded literal. X is a relation over $H \times H$, where $(h_j, h_i) \in X$ means h_j could explain h_i. The relation X is constrained so that the resulting explanation graph is acyclic. Next, $P : H \rightarrow [0, 1]$ is a plausibility function, $S : H \rightarrow \{Accepted, Rejected, Undetermined\}$ gives the belief status of a hypothesis, and $V \subseteq H$ is a set of evidence hypotheses that, when accepted, are considered to require an explanation. I is an irreflexive, symmetric relationship over H where $(h_j, h_i) \in I$ means h_j is incompatible with h_i. The sets I and X are constrained so that $X \cap I = \emptyset$. Additionally, $(\forall (h_j, h_i) \in I)(S(h_j) = Accepted \rightarrow S(h_i) = Rejected)$. We say that if $(h_j, h_i) \in X$ and $S(h_j) = S(h_i) = Accepted$, then h_j explains h_i and h_i is explained by h_j. This use of *explained* and *explained by* does not require or imply that either h_j or h_i is unique in its respective role. An *explanation* is a set of accepted hypotheses, and a *consistent, complete explanation E of V* is an explanation such that (1) for every $h \in E$, no hypothesis incompatible with h is also in E, i.e., E is consistent, and (2) every accepted hypothesis $h_V \in V$ is explained by some hypothesis $h \in E$, i.e., E is complete.

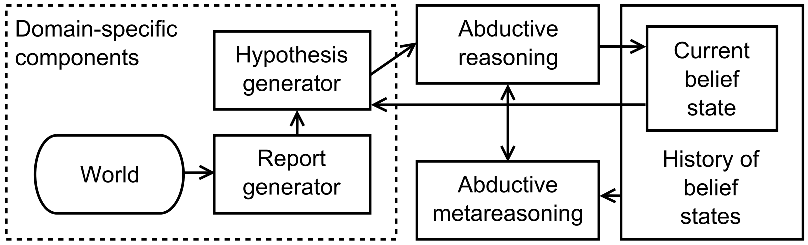

Figure 1. System architecture. Domain-specific components are separate from domain-general reasoning and metareasoning components. Reports, which might be noisy, are obtained from the world. The plausibility of each report is calculated according to domain knowledge and current beliefs. Each report requires explanation, as do any unexplained beliefs (except top-level beliefs). Hypotheses are generated by a domain-specific function and then reviewed (accepted or rejected) by the abductive reasoning procedure. Newly-acquired beliefs might themselves require explanation, and the process starts again. If any reports or beliefs remain unexplained, which we call anomalies, abductive metareasoning is activated in order to determine the causes of the anomalies. Abductive metareasoning might ask the domain-specific components to generate new hypotheses, revise the belief state, and/or leave the anomalies unexplained. Reports that remain unexplained are deemed to be noise.

We do not require that a *could explain* relation $(h_j, h_i) \in X$ be interpretable as material implication or logical entailment, as is sometimes the case in other treatments of abduction (Aliseda, 2006; Kakas, Kowalski, & Toni, 1992). Instead, we suppose that hypotheses represent possible causal relations (h_j is a possible cause of h_i).

In order to experimentally evaluate abductive reasoning and metareasoning across different problem domains, we separate the problem domain from the reasoning system. The system architecture is shown in Figure 1. This leads to another definition.

Definition 3.2. A *problem domain* M provides two functions, OBSERVE and GENERATEHYPOTHESES.

$$\text{OBSERVE}(M, F) = (H_{\text{reports}}, P) \tag{1}$$

The OBSERVE function generates reports H_{reports} of observed properties of the world. These reports come with plausibilities defined by P. The set F may be used to focus the observations on particular features of the world. As used by abductive metareasoning, detailed in Section 5, F is a set of unexplainable reports. When $F = \emptyset$, the OBSERVE function can be thought of a general scan of the world without a particular focus. In an object tracking domain, this general scan might report highly plausible object detections that survive a high-threshold noise filter. In a medical diagnosis domain, this general scan might contain the set of answers to typical questions about the patient's symptoms and history. When $F \neq \emptyset$, it contains unexplainable reports that might require corroborating reports or more detailed reports in order to be explainable. For example, F might contain results from a medical test that, on their own, have no plausible explanation, but might be better explained if results are gathered from a follow-up test.

$$\text{GENERATEHYPOTHESES}(M, H_{\text{unexp}}, H_{\text{acc}}) = (H_{\text{exp}}, X, P, I) \quad (2)$$

The GENERATEHYPOTHESES function produces new hypotheses H_{exp} that purport to explain reports in H_{unexp} such that these hypotheses are consistent with accepted hypotheses H_{acc}. The new explanatory relations, plausibilities of the generated hypotheses, and incompatibility relations are given by X, P, and I, respectively.

Abstractly, we can think of M as containing the knowledge about a problem domain. As far as the abductive reasoning algorithms are concerned, M and its corresponding functions are black boxes. This separation enables us to experiment with different domains without modifying the reasoning engine.

Abductive reasoning, as used in our system, is characterized by additional definitions.

Definition 3.3. A *partial abduction* is an operation on belief states where $\text{PARTIALABDUCE}(B_1) = B_2$ means B_2 is produced from B_1 either by changing nothing ($B_1 = B_2$) or *accepting* one hypothesis and *rejecting* incompatible hypotheses, should any exist. A partial abduction meets four constraints:

1. If no unexplained evidence have hypotheses, or no evidence is unexplained, then the belief state is unchanged.
2. If there exists a hypothesis that is undecided (not accepted or rejected) for some unexplained evidence and is sufficiently plausible to be accepted, then some hypothesis is accepted.
3. One or zero hypotheses are accepted.
4. When a hypothesis is accepted, all incompatible hypotheses are rejected.

Lemma 3.1. If $\text{PARTIALABDUCE}(B_1) = B_2$, then either $\text{UNEXPLAINED}(B_2) \subset \text{UNEXPLAINED}(B_1)$ or $B_1 = B_2$. In other words, the PARTIALABDUCE function, applied to a belief state, reduces the number of unexplained reports or leaves the belief state unchanged.

Proof sketch. Assume $\text{PARTIALABDUCE}(B_1) = B_2$ and suppose $B_1 \neq B_2$. Then some hypothesis h was accepted by the partial abduction operation, and the reports explained by h are no longer unexplained. □

Definition 3.4. A belief state B is *finalized* if $\text{PARTIALABDUCE}(B) = B$.

In these terms, an *abductive reasoning algorithm* is tasked with transforming an initial belief state into a finalized belief state by performing partial abductions. Next, we detail the abductive reasoning algorithm used in our system.

4. The EFLI Algorithm

Bylander et al. (1991) show that certain classes of abduction problems cannot efficiently be solved. We take an efficient greedy approach to the abduction problem, similar to that implemented in Josephson and Josephson's (1994) PEIRCE-IGTT system. This realizes an algorithm called *EFLI*: *Essentials First, Leveraging Incompatibility*, which iteratively accepts one hypothesis and rejects incompatible hypotheses, until either all evidence is explained or no other hypotheses for the unexplained evidence are available or sufficiently plausible. Hypotheses are grouped into *contrast sets*, where each contrast set contains all the plausible hypotheses for some report. Essential hypotheses are accepted first. An essential hypothesis is the sole member of a contrast set, so it is the only plausible hypothesis for some report. Unless it is accepted, some evidence would remain unexplained. Then, hypotheses are ordered for acceptance by the degree to which the best hypothesis in a contrast set surpasses the second best hypothesis, in terms of plausibility. Although not shown here, the steps in the EFLI algorithm satisfy the requirements of a PARTIALABDUCE function.

The EFLI algorithm can be thought of as a specialization of a general abduction algorithm. This general abduction algorithm is given by the pair of functions FINALIZE and ABDUCE (Algorithm 1). The ABDUCE function is responsible for obtaining reports and hypotheses, and adding them to the belief state. Then FINALIZE is called, which iteratively calls PARTIALABDUCE until all reports are explained or there are no more hypotheses that are acceptable. The ABDUCE function is parameterized in part by η, the minimum plausibility threshold for a hypothesis to be considered. This parameter η is taken into account by the ADDHYPOTHESESTOBELIEFSTATE function, which rejects hypotheses (after adding them to the belief state) whose plausibilities are less than η.

Algorithm 1 The FINALIZE and ABDUCE functions.

```
function FINALIZE(B)
    B' ← PARTIALABDUCE(B)              ▷ Possibly accept one hyp., reject incompatible hyps.
    while B' ≠ B do                    ▷ As long as a hypothesis was previously accepted...
        B ← B'
        B' ← PARTIALABDUCE(B)          ▷ Possibly accept another hyp., until abduction is complete
    end while
    return B'
end function
function ABDUCE(M, B_0, η, DoMetareasoning?)
    (H_reports, P) ← OBSERVE(M, ∅)                                ▷ Generate reports
    B_1 ← ADDREPORTSTOBELIEFSTATE(B_0, H_reports, P)              ▷ Save in belief state
    H_unexp ← UNEXPLAINED(B_1)                                    ▷ Gather unexplained evidence
    H_acc ← ACCEPTED(B_1)                                         ▷ Gather accepted hypotheses
    (H_exp, X', P', I') ← GENERATEHYPOTHESES(M, H_unexp, H_acc)   ▷ Gen. new hypotheses
    B_2 ← ADDHYPOTHESESTOBELIEFSTATE(B_1, H_exp, X', P', I', η)   ▷ Save in belief state
    B_3 ← FINALIZE(B_2)                                           ▷ Build an explanation
    if DoMetareasoning? then                  ▷ Check if metareasoning should be attempted
        B_4 ← METAREASON(B_3)                 ▷ Do metareasoning; refer to Section 5
        return B_4                            ▷ Return belief state after metareasoning
    else
        return B_3                            ▷ Return belief state (no metareasoning activated)
    end if
end function
```

Theorem 4.1. The FINALIZE function is guaranteed to terminate.

Proof. From lemma 3.1, we have that either the partial abduction leaves a belief state unchanged, causing termination of the loop, or the partial abduction reduces the set of unexplained evidence. Since the set of hypotheses, which includes the evidence, is finite, it follows that the set of unexplained evidence is finite, and therefore the algorithm is guaranteed to terminate. □

EFLI (Algorithm 2) serves as an implementation of the PARTIALABDUCE function. EFLI uses two relations in order to pick out the best hypotheses:

- $\geq_{\text{contrast}}$, a preference relation on contrast sets.
- $\geq_{\text{hyp}}$, a preference relation on hypotheses in a contrast set.

These relations are defined as,

$$\{h_1, \ldots, h_m\} \geq_{\text{contrast}} \{h'_1, \ldots, h'_n\} \quad \text{iff} \quad P(h_\alpha) - P(h_\beta) \geq P(h'_\alpha) - P(h'_\beta), h \geq_{\text{hyp}} h' \quad \text{iff} \quad P(h) \geq P(h')$$

where h_α and h_β are the two most plausible hypotheses in the contrast set $\{h_1, \ldots, h_m\}$, and h'_α, h'_β likewise for the contrast set $\{h'_1, \ldots, h'_n\}$.

Algorithm 2 The EFLI partial abduction function.

```
function PARTIALABDUCE_EFLI(B_0)
    Φ ← CONTRASTSETS(B_0)
    if Φ = ∅ then                                ▷ No contrast sets, nothing to accept
        return B_0
    else
        Γ ← max_{≥contrast} Φ                    ▷ Find most decisive contrast set
        h ← max_{≥hyp} Γ                         ▷ Find best hypothesis
        B_1 ← ACCEPT(B_0, {h})                   ▷ Accept h
        B_2 ← REJECT(B_1, INCOMPATIBLE(B_0, h))  ▷ Reject incompatible hypotheses
        return B_2
    end if
end function
```

The object tracking domain might be realized in this framework as follows. M contains knowledge of cameras, average object speed, and so on. OBSERVE(M, F) queries the cameras and performs blob detection. These detections are returned as reports (which are grounded literals) with associated plausibility estimates. The plausibility of a report represents the blob detector's confidence that the blob is an object of interest. The set F may contain unexplained reports that may be used to point and zoom the cameras to the regions where these reports originated, in the hope of generating new detections which can shed light on the unexplained reports. The GENERATEHYPOTHESES$(M, H_{\text{unexp}}, H_{\text{acc}})$ function receives unexplained reports and accepted object movement hypotheses (i.e., current beliefs about recent object movements and locations) and generates new object movement hypotheses that can explain the reports but are consistent with current beliefs.

Two object movement hypotheses are inconsistent if they both state that the same object moved in two different directions or two different objects arrived at the same place. We are assuming in this example that objects do not split or merge. In some cases, no hypothesis may be generated for a report if the object cannot possibly have moved as far as the report seems to indicate.

The EFLI abduction algorithm builds a consistent explanation of the reports by accepting one object movement hypothesis at a time and rejecting those that are incompatible. More decisive and plausible hypotheses are preferred first. Ultimately, either all reports are explained or no plausible object movement hypotheses remain because some were rejected due to incompatibility or low plausibility, or never generated to explain certain reports. Next we describe a metareasoning procedure to handle cases of unexplainable reports.

5. Metareasoning

We call *anomalies* those reports and other evidence that remain unexplained in a finalized belief state. The system checks for their presence in a metareasoning function, METAREASON, that is activated from the ABDUCE function (Algorithm 1). In some cases, domain knowledge or background knowledge is insufficient, causing hypotheses not to be generated for some true reports. However, in this work we assume that domain knowledge is sufficient to generate true hypotheses for all true reports, assuming the current world estimate is accurate. Under this assumption, in domains where all reports are guaranteed to be true (noise-free conditions), anomalies are necessarily the result of belief errors. However, in more realistic environments, not all unexplainable reports are true reports; some reports might be unexplainable partly because they are noise and do not warrant any explanation. Part of the challenge of metareasoning is to identify which reports are unexplainable due to errors in the belief state and which are due to false reports (and hence, not due to errors). The metareasoning task can be construed as an abductive one by treating the anomalies as a kind of *meta-evidence* that require explanation by *meta-hypotheses*. Such a metareasoning system is able to use the same abductive reasoning machinery employed by the base-level reasoner.

The remainder of this section details the possible causes of anomalies, their corresponding belief revisions, and the criteria that tell us (as experimenters) whether the revision is correct. Note that an anomaly might have multiple possible causes. The METAREASON function handles the following tasks. For each possible cause, a meta-hypothesis is generated, which specifies the cause, the revision, the subset of anomalies it is said to explain, and an estimated plausibility score. The meta-hypotheses are added to a *meta-belief state* (so that the same abduction algorithms defined above may be employed), and abductive reasoning commences, yielding a set of accepted meta-hypotheses. The belief revisions that are specified by the accepted meta-hypotheses are applied to the original belief state, which is then finalized. If any anomalies remain (or new anomalies appear), metareasoning is activated again on the new belief state. Care is taken not to generate meta-hypotheses that have already been evaluated. This ensures that the procedure halts, although we do not provide a proof here.

In the following discussion of meta-hypotheses, let B be a belief state, $A = \text{ANOMALIES}(B)$, and $H_A = \bigcup_{h \in A} \text{HYPOTHESES}(B, h)$, i.e., the set of hypotheses that could explain some of the anomalies. If any such hypotheses H_A exist, they necessarily were rejected due to incompatibility or low plausibility.

5.1 Implausible Hypotheses

Some reports may be anomalous due to the rejection of one or more implausible hypotheses. These rejected hypotheses are characterized by $H_P = \{h | h \in H_A \wedge P(h) < \eta\}$, where P is the plausibility function of B and η is the minimum plausibility threshold. Unrejecting one or more of these implausible hypotheses, thus possibly allowing their acceptance, might eliminate some anomalies.

For each $h \in H_P$, we hypothesize that the rejection of h is responsible for some reports having no explanation. The plausibility is estimated by $p = \left[P(h) + \sum_{r \in R} P(r)\right] / (\|R\| + 1)$, where $R \subseteq A$ contains the anomalies that are eliminated by applying the revision and finalizing the resulting belief state. This estimate is higher when the plausibility of h and what it explains are higher. We have found that this plausibility estimate works reasonably well in empirical studies. For evaluating experimental results, we stipulate that the revision is correct if $R \neq \emptyset$, a greater number of the eliminated anomalies in R are true evidence rather than false, and the hypothesis h that is unrejected is a true hypothesis.

5.2 Incompatible Hypotheses

Some reports may be anomalous due to some of the hypotheses being rejected upon the acceptance of other hypotheses. Let $H_I = \{h | h \in \textsc{Accepted}(B) \wedge \textsc{Incompatible}(B, h) \cap H_A \neq \emptyset\}$. The set H_A contains all hypotheses that could explain some of the anomalies. The set H_I contains accepted hypotheses that are incompatible with members of H_A. An accepted hypothesis $h \in H_I$ may have been responsible for rejecting some hypotheses; thus, if the status of h is changed to undecided and then rejected (to prevent it from being accepted again), some anomalies might be eliminated.

For each $h \in H_I$, we hypothesize that the acceptance of h is responsible for some reports having no explanation. The set $R \subseteq A$ contains the anomalies that actually are eliminated by rejecting h. The plausibility is estimated by $p = (1 - P(h)) * \left(\sum_{r \in R} P(r) / \|R\|\right)$. This estimate is higher when h is less plausible and the anomalous reports are more plausible. We count the revision as correct if $R \neq \emptyset$, a greater number of the eliminated anomalies in R are true than false, the hypothesis h that is to be rejected is false, and some hypothesis that h had precluded is in fact true.

5.3 Insufficient Evidence

Some anomalous reports might have no hypotheses at the base level, so the first two scenarios are inapplicable. An anomaly may have no hypotheses for one of two reasons: either prior accepted hypotheses limit possible hypotheses for the new reports to the point that none are offered, or there are insufficient reports to reasonably narrow the set of hypotheses. This latter case will be described first.

Any reasonable implementation of the function GENERATEHYPOTHESES will generate only those hypotheses that can meaningfully explain the reports. Given a dearth of evidence, one would expect a reasonable problem domain not to generate an infinite (or very large) set of overly-specific hypotheses. For example, in a medical diagnosis domain, the single report "headache" might be explainable by any one of (or combinations of) many diseases and ailments. But no bounded rational agent would methodically consider each of these hypotheses. Instead, more evidence would be gathered.

The *insufficient evidence* meta-hypothesis supposes that some anomaly is unexplainable because no hypotheses were offered due to insufficient evidence. Thus, the corresponding revision requires first seeking more evidence, preferably reports that corroborate or add detail to the anomalous report in question, and then generating new hypotheses. The plausibility of such a meta-hypothesis is estimated by $p = P(r)$, where r is the anomaly that *might* be eliminated by gathering more evidence. We count the revision as correct if r is true.

Whether or not such a revision is effective (more reports are obtained and new hypotheses for some anomalies are generated and accepted) is not known until the meta-hypothesis is accepted during abductive metareasoning and the revision is applied. This conservative approach is taken because sometimes obtaining more evidence (e.g., performing medical tests) is costly and/or harmful and should be performed only if other meta-hypotheses are ruled out.

5.4 Order Dependency

The final possible cause of an anomaly is that no hypotheses were ever offered due to the cognitive system mistakenly believing that the report is unexplainable (i.e., impossible) given its current world estimate. We suppose that the problem domain's GENERATEHYPOTHESES function is defined to generate only those hypotheses that are consistent with the current belief state. Thus, if some of the accepted hypotheses in the belief state (which were accepted to explain earlier reports) are false, then the GENERATEHYPOTHESES function might fail to generate hypotheses for true reports.

The *order dependency* meta-hypothesis suggests that one or more anomalies have no hypotheses because prior accepted hypotheses were in error, and that they should be reconsidered *in light of* recently-obtained reports. In other words, the anomalies are the result of the particular order the reports were obtained. The revision involves identifying a previous belief state to revert to (and thereby erasing recently generated and accepted hypotheses), then injecting recent reports, generating new hypotheses (given the less committed belief state and more reports), and finalizing the belief state. It is not clear, at the time of this writing, how far back the belief state must be reverted. In the experiments detailed here, the system reverts to the belief state immediately preceding the introduction of the reports that ultimately proved to be anomalous. The plausibility of such a meta-hypothesis is estimated as $p = \sum_{r \in R} P(r)/\|R\|$, where $R \subseteq A$ is the set of anomalies that *might* be eliminated by reconsidering previously-accepted hypotheses in light of subsequent reports. We say that the revision is correct if more of the eliminated anomalies in R are true than false. Like the insufficient evidence meta-hypothesis, an order dependency revision is not attempted until the corresponding meta-hypothesis is accepted as part of an explanation of some anomalies. This is because generating new hypotheses in light of subsequent reports might be costly.

5.5 Noise Detection

Not all anomalies should trigger belief revisions. Some reports might be unexplainable because they are false, i.e., noisy reports. Though some of these noisy reports might be explainable by accepting implausible but false hypotheses, for example, the correct action is to reject the anomalous reports so that they are no longer considered unexplained evidence. This is achieved by treating the noise hypothesis as a fallback meta-hypothesis when no other meta-hypothesis is sufficiently plausible. We find that a minimum plausibility threshold for abductive metareasoning, η_{meta}, is effective for filtering out implausible meta-hypotheses. Anomalies that remain unexplained after abductive metareasoning are considered to be noise, and in turn rejected. Because these reports are rejected, they no longer manifest as anomalies.

The next two sections detail our experimental evaluation of both the base-level abductive reasoning system and the combined abductive reasoning and metareasoning system.

6. Experimental Evaluation

We experimented with two object tracking domains, one simulated and one based on aerial video surveillance. Although we believe that many different kinds of intelligent systems tasked with different problem solving goals can benefit from an abductive metareasoning system, object tracking tasks expose the usefulness of abductive metareasoning in the following ways.

- The task is easily framed in explicitly abductive terms, in which object detections make up reports and object movements serve as the hypotheses.
- As will be shown, it is practical to establish a minimum plausibility for movement hypotheses, though anomalies due to implausible hypotheses might result.
- In the simulated tracking domain, movement hypotheses are incompatible if they describe the same object in two different locations at the same time. Thus, anomalies resulting from incompatible hypotheses are possible.
- It is often practical for surveillance systems to filter out sensor detections that do not meet a minimum threshold of plausibility in an attempt to filter out noise. In some cases, however, this filtering causes anomalies due to insufficient evidence.
- Future movement hypotheses depend on the system's current estimate of the situation, i.e. its beliefs, which are formed from the acceptance of prior movement hypotheses. Consequently, false beliefs might cause order dependency anomalies.

In the simulated tracking domain, the cognitive system obtains reports about moving objects in a 10×10 discrete grid. This grid constitutes the world, and is fully observable, with one caveat described below. The objects' movements are random walks. At each time step, each object makes a fixed number of random one-step movements, which we call *grid steps* (diagonals not allowed). Virtual sensors report the final location of each object's walk in that time step. No two objects are allowed to occupy the same grid cell at the completion of their walk. Thus, there is no need to handle merges and splits.

For simplicity, each object bears a unique color. An object's color is a stand-in for any variety of more realistic object features that support its identification. However, the center 50% of the grid is watched only by sensors that do not detect color. All objects in that area are seen as gray, and are therefore indistinguishable. The outer 50% is watched by sensors that do report color. When

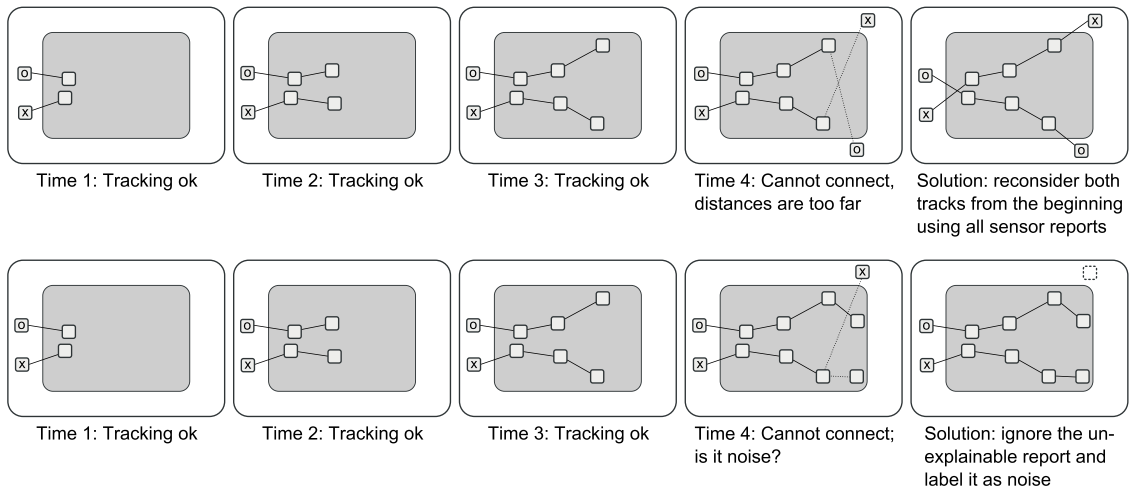

Figure 2. Anomalies due to order dependency (top) and noise (bottom) in the simulated object tracking domain. Reports for objects in the middle gray area do not specify the object's color (here represented by 'x' and 'o'), thus any object within a short range could possibly account for the detection.

objects move into this outer area, they can be identified. The presence of this gray area introduces the possibility of anomalies due to mis-tracking. The upper diagrams of Figure 2 illustrate an anomaly of this sort. Noisy reports are simulated by introducing fake reports which describe non-existent objects and by distorting (randomly modifying) reports about actual objects. Each object report has a certain chance of distortion, which in our experiments ranged from 0% to 30%. The lower diagrams of Figure 2 illustrate an anomaly due to noise.

Reports provided by the OBSERVE function take the form "an object was detected at location x, y at time t with color c (or gray)." Reports are assigned random plausibility scores such that noisy reports typically score low and true reports score high. These plausibility scores simulate scores produced by blob detection schemes and similar processes in actual tracking systems. A threshold is established so that an initial call to the OBSERVE function returns only highly plausible reports, but additional calls (focusing on certain anomalies) return all nearby reports regardless of their plausibility. Hypotheses generated by the GENERATEHYPOTHESES function take the form "The object with color c moved from x, y at time t to x', y' at time t'." Two movement hypotheses are incompatible if they claim that two objects moved into the same location or out of the same location (the domain does not handle merges and splits), or that the same object (identified by color) is in two different locations at the same time. Movement hypotheses are scored based on the distance of the movement. The system is provided 1,000 examples of real object movements and builds a model of the probability of movements of various distances. This simple model is used to estimate the plausibility of movement hypotheses.

In addition to the simulated object tracking domain, we experimented with object tracking from aerial imagery using the KIT AIS Data Set[1] (Schmidt & Hinz, 2011). The task is the same as in

1. `http://www.ipf.kit.edu/downloads_People_Tracking.php`

simulated object tracking: to infer object (person) movements. In each frame, hundreds of very small person-like objects are detected. We take the detections as reports and perform object tracking in the same manner as the simulated tracking domain. Each detection is assigned a plausibility by the detector, and the OBSERVE function initially provides only highly plausible detections. If queried for more observations nearby to some anomalous reports, the threshold limitation is eliminated for reports in that region. Note that no pair of hypotheses in this domain are incompatible with each other (merges and splits are possible).

Three metareasoning strategies are compared: *abd-estimate*, as described in Section 5; *ignore*, which simply rejects all anomalies, effectively labeling them as noise; and *oracle*, which is the same abductive metareasoning strategy as *abd-estimate* but where the plausibilities of meta-hypotheses are set so that true meta-hypotheses have score 1.0, false have score 0.0, and $\eta_{\text{meta}} = 0.01$ to ensure that only true meta-hypotheses are accepted. Oracle metareasoning performance is the maximum performance that abductive metareasoning can be expected to achieve; however, it does not guarantee perfect performance because some errors and noise never manifest as anomalies.

Three distinct hypotheses guided these experimental investigations:

Hypothesis I: Abductive reasoning with abductive metareasoning gives good performance in terms of precision, recall, and noise detection metrics in both domains. Furthermore, best performance is achieved when the minimum plausibility $\eta > 0$, thus demonstrating the utility of screening out low plausibility hypotheses.

Hypothesis II: Metareasoning gives better performance on these metrics compared with no metareasoning. We also expect that oracle metareasoning consistently performs best.

Hypothesis III: Each of the four types of meta-hypotheses are accepted, at least in some cases, indicating that each kind of meta-hypothesis has a useful role.

We performed 30 random simulated tracking scenarios for each minimum plausibility η value and a single experiment for each η value with the aerial tracking dataset. In simulated object tracking, each simulation has ten time steps and six different objects moving about. At each time step, each object took a random walk of six grid steps. In either domain, when metareasoning is enabled, some reports might also be rejected as noise, which we call *noise claims*. The levels of noise in both domains are very high. In the simulated domain, about 25% of reports are false. In the aerial domain, about 73% of reports are false according to the ground truth established in the data set.

We define four metrics to gauge the accuracy of the cognitive system's explanations and noise detection:

$$\begin{aligned}
\text{Precision} &= \frac{\|\text{Actual movements} \cap \text{Accepted movement hypotheses}\|}{\|\text{Accepted movement hypotheses}\|} \\
\text{Recall} &= \frac{\|\text{Actual movements} \cap \text{Accepted movement hypotheses}\|}{\|\text{Actual movements}\|} \\
\text{Noise precision} &= \frac{\|\text{Actual noisy reports} \cap \text{Noise claims}\|}{\|\text{Noise claims}\|} \\
\text{Noise recall} &= \frac{\|\text{Actual noisy reports} \cap \text{Noise claims}\|}{\|\text{Actual noisy reports}\|}
\end{aligned}$$

Table 1. **Simulated tracking.** Performance in noise-free (N % = 0) and noisy conditions (N % = 30), $\eta = 0.10, \eta_{\text{meta}} = 0.40$, supporting Hypotheses I and II. Under noise-free conditions, noise precision and noise recall are necessarily 0. Asterisks in *abd-estimate* scores indicate the difference with *ignore* scores is statistically significant (* $p < 0.05$, ** $p < 0.01$).

N %	Metareasoning	Precision	Recall	Noise Precision	Noise Recall
0	abd-estimate	0.875 ± 0.021	0.777 ± 0.022 *		
0	ignore	0.862 ± 0.023	0.715 ± 0.032		
0	oracle	0.916 ± 0.013	0.843 ± 0.016		
30	abd-estimate	0.664 ± 0.017 **	0.526 ± 0.014 **	0.588 ± 0.012 **	0.709 ± 0.015
30	ignore	0.634 ± 0.018	0.481 ± 0.016	0.537 ± 0.011	0.715 ± 0.013
30	oracle	0.767 ± 0.015	0.623 ± 0.013	0.679 ± 0.011	0.780 ± 0.012

Results are shown in Table 1 and Figure 3 for the simulated tracking domain and Table 2 for the aerial tracking domain. For the simulated tracking domain, each of the three kinds of metareasoning that we compared exhibited the best trade off between precision and recall at $\eta = 0.10$, where this trade off is calculated as the harmonic mean of the two metrics (often called the F1 measure). Table 1 shows the metrics at $\eta = 0.10$. Best performance in the aerial domain for abductive reasoning (without metareasoning) was found at $\eta = 0.80$. At this same parameter value, abductive metareasoning brought the most benefit. Results are shown in Table 2.

Performance in simulated tracking is low in noisy conditions, but this is to be expected due to the inherent difficulty of the task. The object movements are random and in half of the grid the objects are indistinguishable. Furthermore, some reports are noisy. Table 1 shows performance in both noise-free and noisy conditions. Noise-free conditions yield considerably higher performance.

In summary, Hypothesis I is supported by the evidence. Furthermore, Table 1 and Table 2, for simulated tracking and aerial experiments, respectively, show that *abd-estimate* metareasoning is better than *ignore* metareasoning, at the right η values, and worse than *oracle* metareasoning. For the simulated tracking domain, in which we executed multiple experiments with different random variations, we see that in noisy conditions, *abd-estimate* performs significantly better in terms of precision, recall, and noise precision. Thus, Hypothesis II is also supported.

Finally, Hypothesis III is supported by evidence from experiments with the simulated tracking domain, which show that each meta-hypothesis played a role. At $\eta = 0.10$, on average 1.68 ± 0.27 (standard error) *implausible hypotheses* meta-hypotheses were accepted during each experiment, 4.68 ± 0.29 *incompatible hypotheses*, 8.56 ± 0.72 *insufficient evidence*, and 3.42 ± 0.32 *order dependency* meta-hypotheses. In the aerial domain, we find that six *implausible hypotheses* and two *insufficient evidence* meta-hypotheses were accepted at $\eta = 0.80$. *Order dependency* and *incompatible hypotheses* meta-hypotheses were never considered to be possible causes of anomalies because there are no incompatible pairs among the hypotheses in this domain.

7. Related Work

The present work combines abductive reasoning and metareasoning. Few cognitive systems have taken this approach. A notable exception is work by Bharathan (2010), which explored metarea-

Table 2. **Aerial tracking.** Performance in noisy conditions (most reports are spurious and not reports of actual person detections), $\eta = 0.80, \eta_{\text{meta}} = 0.60$, supporting Hypotheses I and II. The original dataset was not modified to add or remove varying degrees of noise.

Metareasoning	Precision	Recall	Noise Precision	Noise Recall
abd-estimate	0.660	0.825	0.972	0.853
ignore	0.636	0.700	0.911	0.853
oracle	0.680	0.850	0.972	0.853

soning as applied to an abductive reasoning system. However, the metareasoning facility there is somewhat *ad hoc* in its design and does not benefit from the architectural simplicity of our abductive metareasoning architecture. Additionally, the system only considers order dependency meta-hypotheses and does not attempt to detect noise. The present system was built partly to bring a unified architecture to abductive reasoning and metareasoning and to support a greater variety of meta-hypotheses.

In the remainder of this review, we will first give a brief overview of abductive reasoning, then look at metareasoning. Computational approaches to abductive reasoning can be divided roughly into three styles. Pagnucco (1996) distinguishes between two styles, *logic-based* and *set-covering*. We add to these *probabilistic* abduction, and give a short description of each. Our approach, characterized by the *EFLI* algorithm, most closely matches the set-covering style. A more thorough account of the various approaches to abduction is given by Schurz (2008).

Logic-based abduction typically reifies the concept p *explains* q as $\Theta \cup \{p\} \vdash q$, where Θ is the background theory and $\Theta \cup \{p\}$ is consistent. The task is to find p given that q has been reported and requires explanation (i.e., $\Theta \nvdash q$). One kind of logic-based abduction is Abductive Logic Programming (ALP), introduced by Kakas et al. (1992) and directly influenced by earlier work on THEORIST (Poole, Goebel, & Aleliunas, 1987). ALP has been implemented as an extension to Prolog (Fung & Kowalski, 1997) and subsequently integrated with constraint programming (Kakas, Michael, & Mourlas, 2000; Endriss et al., 2004). Another approach to logic-based abduction that attempts to find an explanation p to explain (logically entail) q makes use of semantic tableaux (Aliseda, 2006). These approaches typically seek simpler explanations rather than ranking possible explanations by some non-logical metric such as plausibility. Explanation specificity may serve as an additional criterion, though the degree of specificity an explanation should possess is a trade-off between informativeness and correctness (Hobbs et al., 1990). Ng and Mooney (1990) add *explanatory coherence* as another feature of good explanations. Coherent explanations are explanations that tie together or make sense of many observations.

An alternative kind of logic-based abduction is typified by AbRA (Bridewell & Langley, 2011), which iteratively builds hierarchical explanations by instantiating rules with observed constants or defaults. This is in contrast to the EFLI algorithm, which assumes reports and hypotheses are grounded literals, and therefore requires all reports and hypotheses to be generated ahead of time and available before abductive reasoning begins building an explanation. AbRA, on the other hand, employs heuristics for selecting which reports to attempt to explain and which rules to instantiate first. In this way, it is able to operate with larger sets of reports and possible hypotheses.

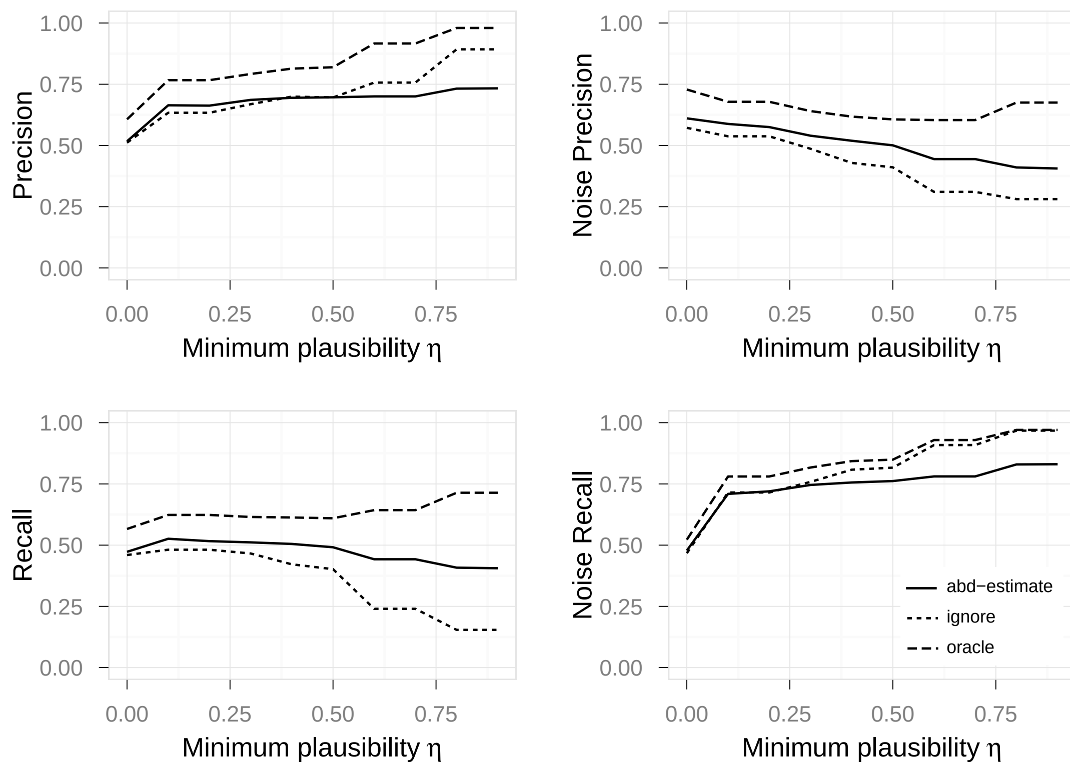

Figure 3. **Simulated tracking.** Average performance across 30 random cases in the simulated object tracking domain for various minimum plausibility η values and $\eta_{\text{meta}} = 0.40$, supporting Hypotheses I and II.

UMBRA (Meadows, Langley, & Emery, 2013) follows a similar process but uses somewhat different heuristics from its predecessor.

The EFLI algorithm realizes a set-covering pattern of abduction. Set-covering abduction operates on set of effects, a set of causes, and a relationship between the two sets that specifies which causes can possibly explain which effects (as in our relation X). The task is to find a subset of the possible causes that is both internally consistent (some causes might be incompatible with each other, as in our relation I), minimal in some sense, and explains all of the effects. Reggia et al. (1983) give an early example of this approach, although in that work, no possible causes were incompatible. It is clear that *EFLI* is a kind of set-covering abduction algorithm, though it is more directly influenced by the *hypothesis assembly* approach of Bylander et al. (1991). However, *EFLI* adds *pragmatic* concerns by preferring hypotheses that more significantly surpass their rivals in terms of plausibility, and by establishing a minimum plausibility threshold η.

Probabilistic abduction involves finding the most probable set of propositions (one from each variable of interest) given some evidence. This set of propositions is said to explain the evidence (Pearl, 1987). Pearl refers to this process as both belief revision and abductive inference (Pearl, 1994). Poole (1993) connects probabilistic abduction and logic-based abduction. Our research has focused on categorical beliefs, and some of these may turn out to be false. The role of the metareasoning system is to discover these false beliefs and root them out. When beliefs are only probabilistic, a completely different kind of metareasoning system should be devised. We have not yet explored this area of research.

The ideas of *metareasoning*, *meta-level knowledge*, *meta-rules*, and so on have been explored for much of the history of the field of Artificial Intelligence. Summaries of recent developments in metareasoning, or *metacognition* in general, include reviews by Cox (2005) and Anderson and Oates (2007). These reviews show that metacognition may be found in mathematics, psychology, artificial intelligence, and philosophy. Its varieties are too broad and numerous to cover here, but we note that metacognition and metareasoning are not only of interest to artificial intelligence researchers.

Anomaly-driven theory revision has been explored previously. An early system by Karp (1989) responds to unexplained experimental outcomes, i.e., prediction failures, by designing modifications to the theory that, when applied, produce a theory that is able to predict the observed outcome. Bridewell (2004) describes a system with a similar goal. Bridewell's work maintains the assumption that reports from the world are noise-free. Karp's method, on the other hand, might simply fail to produce an acceptable theory revision. In this sense, it is similar to how abductive metareasoning decides whether or not a report is noisy.

The Meta-Cognitive Loop (MCL) from Schmill et al. (2011) shares similarities with the present work. The MCL is a component that attaches to a host reasoning system and is informed by the host system about possible actions and expectations regarding the results of those actions. Then, *in situ*, it monitors the host system's actions and detects expectation violations, which are called anomalies. Causes of the anomalies, and appropriate responses, are determined by consulting domain-general ontologies, represented as Bayesian networks. The present work differs from the Meta-Cognitive Loop component in that, in abductive metareasoning, the variety of possible causes of anomalies is significantly smaller than those represented in MCL's ontologies. The likelihood of each kind of anomaly must be learned in their system, while in the present work, the plausibility of meta-hypotheses are estimated according to domain-general features. Additionally, abductive metareasoning detects noise by way of a generic fallback meta-hypothesis rather than domain-specific noise detectors.

8. Conclusion

This paper has described and evaluated a system for two-level abductive reasoning and metareasoning. It has also shown at least one way to determine correct belief revisions in the face of anomalies, by detailing possible causes of anomalies and estimates of their plausibilities. Abductive metareasoning has proved to be very effective at correcting errors in the belief state and detecting (and ignoring) noisy reports. This effectiveness has been demonstrated in simulated and aerial object tracking tasks, both of which were challenging due to limited sensor capabilities (such as inability to detect color in the simulated domain and very low resolution in the aerial domain) and very high levels of noise.

Although not discussed here, we have evaluated abductive metareasoning in a different world estimation domain and noted similar benefits (Eckroth & Josephson, 2014). In that domain, the cognitive system attempts to explain reports by consulting a Bayesian network world model to find hypotheses and their plausibilities. Although the problem domain is quite different from the object tracking domains presented here, the abductive reasoning and metareasoning components do not require any modifications to handle the task. The only domain-specific parameters that tune the abductive reasoning and metareasoning algorithms are η and η_{meta}. A deeper exploration of each of these domains has been recently completed (Eckroth, 2014).

In future work, we aim to explore a wider variety of problem domains to determine the limits of the domain-generality of this architecture and implementation. One such domain is plan recognition. This domain is a paradigmatic case of abductive reasoning. The task is to observe actions committed by one or more agents and infer the plan that the agents are following (and hence infer their goals). Under normal circumstances, most of the intermediate actions required to execute a plan are hidden from the observer; the observer cannot be sure the agents performed all the steps of the plan. There are usually more than one plausible plans that the agents may be following at any particular time. The observed agents may carry out actions over time, so the most plausible plan given the observed actions may change over time.

It is clear that plan recognition has many of the properties that might make abductive reasoning an effective strategy. Previous work has explored various kinds of abductive reasoning procedures for plan recognition (Paul, 1993; Goldman, Geib, & Miller, 1999). To our knowledge, metareasoning has never been utilized in an abductive plan recognition system. We suspect that metareasoning will lead to improved plan recognition because it is capable of detecting and repairing anomalies, e.g., in cases where observations begin to deviate from the believed plan.

Acknowledgements

We would like to thank the anonymous reviewers of this work for their helpful feedback, as well as Pat Langley, Bruce G. Buchanan, and Reid G. Smith for their detailed comments.

References

Aliseda, A. (2006). *Abductive reasoning: Logical investigations into discovery and explanation*. Norwell, MA: Kluwer Academic Publishers.

Anderson, M. L., & Oates, T. (2007). A review of recent research in metareasoning and metalearning. *AI Magazine*, *28*, 12.

Bharathan, V. (2010). Belief revision in dynamic abducers through meta-abduction. Master's thesis, Department of Computer Science and Engineering, The Ohio State University, Columbus, OH.

Bridewell, W. (2004). *Science as an anomaly-driven enterprise: A computational approach to generating acceptable theory revisions in the face of anomalous data*. Doctoral dissertation, University of Pittsburgh, Pittsburgh, PA.

Bridewell, W., & Langley, P. (2011). A computational account of everyday abductive inference. *Proceedings of the Thirty-Third Annual Meeting of the Cognitive Science Society* (pp. 2289–2294). Boston, MA: Cognitive Science Society.

Bylander, T., Allemang, D., Tanner, M., & Josephson, J. (1991). The computational complexity of abduction. *Artificial Intelligence*, *49*, 25–60.

Cox, M., & Raja, A. (2011). Metareasoning: An introduction. In M. T. Cox & A. Raja (Eds.), *Metareasoning: Thinking about thinking*. Cambridge, MA: MIT Press.

Cox, M., & Ram, A. (1999). Introspective multistrategy learning: On the construction of learning strategies. *Artificial Intelligence*, *112*, 1–55.

Cox, M. T. (2005). Metacognition in computation: A selected research review. *Artificial Intelligence*, *169*, 104–141.

Eckroth, J. (2014). *Anomaly-driven belief revision by abductive metareasoning*. Doctoral dissertation, Department of Computer Science and Engineering, The Ohio State University, Columbus, OH.

Eckroth, J., & Josephson, J. R. (2014). Commonsense abductive reasoning and metareasoning using knowledge from Bayesian networks. *Proceedings of the AAAI-14 Spring Symposium, Knowledge Representation and Reasoning in Robotics* (pp. 31–38). Stanford, CA: AAAI Press.

Endriss, U., Mancarella, P., Sadri, F., Terreni, G., & Toni, F. (2004). The CIFF proof procedure for abductive logic programming with constraints. *Proceedings of the Ninth European Conference on Logics in Artificial Intelligence* (pp. 31–43). Lisbon, Portugal: Springer.

Fung, T., & Kowalski, R. (1997). The IFF proof procedure for abductive logic programming. *The Journal of Logic Programming*, *33*, 151–165.

Goldman, R., Geib, C., & Miller, C. (1999). A new model of plan recognition. *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence* (pp. 245–254). San Francisco: Morgan Kaufmann Publishers.

Hobbs, J. R., Stickel, M., Appelt, D., & Martin, P. (1990). Interpretation as abduction. *Artificial Intelligence*, *63*, 69–142.

Josephson, J. R., & Josephson, S. G. (1994). *Abductive inference: Computation, philosophy, technology*. Cambridge, UK: Cambridge University Press.

Kakas, A., Kowalski, R., & Toni, F. (1992). Abductive logic programming. *Journal of Logic and Computation*, *2*, 719–796.

Kakas, A. C., Michael, A., & Mourlas, C. (2000). ACLP: Abductive constraint logic programming. *Journal of Logic Programming*, *44*, 129–177.

Karp, P. D. (1989). *Hypothesis formation and qualitative reasoning in molecular biology*. Doctoral dissertation, Department of Computer Science, Stanford University, Stanford, CA.

Meadows, B., Langley, P., & Emery, M. (2013). Understanding social interactions using incremental abductive inference. *Proceedings of the Second Annual Conference on Advances in Cognitive Systems* (pp. 39–56). Baltimore, MD: Cognitive Systems Foundation.

Ng, H. T., & Mooney, R. J. (1990). On the role of coherence in abductive explanation. *Proceedings of the Eighth National Conference on Artificial Intelligence* (pp. 337–342). Boston, MA: AAAI Press.

Pagnucco, M. (1996). *The role of abductive reasoning within the process of belief revision*. Doctoral dissertation, Basser Department of Computer Science, University of Sydney, Sydney.

Paul, G. (1993). Approaches to abductive reasoning: An overview. *Artificial Intelligence Review*, *7*, 109–152.

Pearl, J. (1987). Distributed revision of composite beliefs. *Artificial Intelligence*, *33*, 173–215.

Pearl, J. (1994). Belief networks revisited. In D. Bobrow (Ed.), *Artificial intelligence in perspective*, 49–56. Cambridge, MA: MIT Press.

Poole, D. (1993). Probabilistic Horn abduction and Bayesian networks. *Artificial Intelligence*, *64*, 81–129.

Poole, D., Goebel, R., & Aleliunas, R. (1987). Theorist: A logical reasoning system for defaults and diagnosis. In N. Cercone & G. McCalla (Eds.), *The knowledge frontier: Essays in the representation of knowledge*, 331–352. New York: Springer Varlag.

Reggia, J., Nau, D., & Wang, P. (1983). Diagnostic expert systems based on a set covering model. *International Journal of Man-Machine Studies*, *19*, 437–460.

Schmidt, F., & Hinz, S. (2011). A scheme for the detection and tracking of people tuned for aerial image sequences. In U. Stilla, F. Rottensteiner, H. Mayer, B. Jutzi, & M. Butenuth (Eds.), *Photogrammetric image analysis*, 257–270. Berlin: Springer.

Schmill, M. D., Anderson, M. L., Fults, S., Josyula, D., Oates, T., Perlis, D., Shahri, H., Wilson, S., & Wright, D. (2011). The metacognitive loop and reasoning about anomalies. In M. T. Cox & A. Raja (Eds.), *Metareasoning: Thinking about thinking*. Cambridge, MA: MIT Press.

Schurz, G. (2008). Patterns of abduction. *Synthese*, *164*, 201–234.