Biologically Inspired Cognitive Architectures (2014) 7, 53–75

Available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/bica

RESEARCH ARTICLE

# Meta-reasoning for predictive error correction: Additional results from abstraction networks with empirical verification procedures

Joshua K. Jones *, Ashok K. Goel

*Design & Intelligence Laboratory, School of Interactive Computing, Georgia Institute of Technology, United States*

Received 28 August 2013; received in revised form 5 November 2013; accepted 5 November 2013

**KEYWORDS**
Meta-reasoning;
Error-correcting agents;
Self-diagnosis and self-repair;
Perceptual grounding;
Symbol grounding;
Classification learning

**Abstract**
In Jones and Goel (2012), we describe a meta-reasoning architecture that uses abstraction networks (ANs) and empirical verification procedures (EVPs) to ground self-diagnosis and self-repair of domain knowledge in perception. In particular, we showed that when a hierarchical classifier organized as an AN makes an incorrect prediction, then meta-reasoning can help diagnose and repair the semantics of the concepts in the network. Further, we demonstrated that if an EVP associated with each concept in the network can verify the semantics of that concept at diagnosis time, then the meta-reasoner can perform knowledge diagnosis and repair tractably. In this article, we report on three additional results on the use of perceptually grounded meta-reasoning for correcting prediction errors. Firstly, a new theoretical analysis indicates that the meta-reasoning diagnostic procedure is optimal and establishes the knowledge conditions under which the learning converges. Secondly, an empirical study indicates that the EVPs themselves can be adapted through refining the conceptual semantics. Thirdly, another empirical study shows that if EVPs cannot be defined for all concepts in a hierarchy, the computational technique degrades gracefully. While the theoretical analysis provides a deeper explanation of the sources of power in ANs, the two empirical studies demonstrate ways in which the strong assumptions made by ANs in their most basic form can be relaxed.


## Introduction

Recent developments in cognitive neuroscience suggest that brains are fundamentally predictive in nature (Bubic, Von

* Corresponding author.
*E-mail addresses:* jkj@cc.gatech.edu (J.K. Jones), goel@cc.gatech.edu (A.K. Goel).

2212-683X/$ - see front matter 
http://dx.doi.org/10.1016/j.bica.2013.11.003

Cramon, & Schubotz, 2010; Kveraga, Ghuman, & Bar, 2007). Instead of passively receiving information from the external world for internal processing, brains proactively use knowledge available internally to make predictions about the external world, verifying the predictions, and making corrections to their knowledge if and when the predictions turn out to be incorrect. Clark (2012) provides a good review of this research on brains as predictive, error-correcting systems from a cognitive science perspective. As he notes, much of the research on the predictive nature of brains so far has focused on perception and action. Rao and Ballard (1999), for example, provide a computational model of predictive coding in the visual cortex. Kveraga et al. (2007) suggest that the brain's predictive mechanisms are top-down. Bubic et al. (2010) suggest that the predictive brain is the basis of not only of perception, but also of much of cognition. Dietrich (2004) proposes that the predictive brain is also the basis of higher-level cognitive processes of creativity.

In AI too predictions have been the source of power of many a technique for reasoning and learning. As an example, expectation generation is fundamental to Schank's cognitively inspired techniques for sentence, discourse and story understanding (Schank, 1983; Schank & Abelson, 1977). As another example, Winston's (1992) textbook describes identification and correction of mistakes in domain knowledge as a basic learning strategy.

Modern AI often is characterized by the notion of intelligent agents (Russell & Norvig, 2010). An intelligent agent is situated in the external world, with learning and reasoning grounded in perception and action (Wooldridge & Jennings, 1995). We may view an intelligent agent as operating in several ''mental spaces'' (Cox & Raja, 2011; Goel & Jones, 2011, chap. 10): At the ''ground level,'' the agent may map percepts in the world to actions on it; at the ''object level'', the agent may use memory and knowledge to make plans for acting on the world; and at the ''meta-level'', the agent may use meta-reasoning — reasoning about reasoning — to monitor and control its decision making at the object level. This paper is concerned about an agent's use of meta-reasoning for correcting predictive errors by correcting the agent's domain knowledge. A central thesis of this work is that it is useful to ground meta-reasoning in predictive agents in perception and action for the purposes of error correction.

In most agent architectures that include meta-reasoning Cox and Raja (2011), while deliberation at the object level is grounded in perception and action, introspection and reflection at the meta-level typically operate only on the deliberation at the object level, separately from perception and action at the ground level. In contrast, in (Jones & Goel, 2012), we described a computational architecture for grounding meta-reasoning in perception and action. Fig. 1 illustrates our agent architecture with perceptually grounded meta-reasoning. As in traditional agent architectures, meta-reasoning monitors the processing at the object-level and controls it if needed. However, unlike the traditional agent architectures, meta-reasoning in our agent architecture also receives perceptual inputs from the world and selects actions for execution in the world as indicated in Fig. 1.

As we described in (Jones & Goel, 2012), the need for this grounding of meta-reasoning in perception and action arises because of the predictive nature of our intelligent agents. To illustrate, let us consider *compositional classification* in which raw state features of the world are progressively abstracted through a series of classifications in an abstraction hierarchy until a top level target classification is produced. We call this hierarchy an abstraction network (AN). Fig. 2 illustrates a generic AN. Now consider a scenario in which the compositional classifier organized in an abstraction network generates a prediction that at some later time turns out to be incorrect. In this scenario, the meta-reasoner may seek to diagnose and repair the AN. This self-diagnosis and self-repair of the AN entails detection of errors in the processing and/or the knowledge in the AN based on violations of expectations of the environment. In our technique for perceptually grounded meta-reasoning, these expectations are explicitly represented in ANs as *Empirical Verification Procedures* (EVPs), which tie domain knowledge stored at an node in the AN to predictions about the outcomes of actions and observations in the environment. Thus, to employ these EVPs for diagnosis and repair of domain, the meta-reasoner needs to observe not only the processing at the object level, but also the percepts at ground level. Further, when the meta-reasoner identifies problems at the object level through this kind of monitoring, it may need the agent to take some actions in the environment in order to gather more information to resolve the problems. For example, the meta-level may execute EVPs at intermediate nodes in an AN to determine which chunks of knowledge are responsible for an observed top-level prediction error. Finally, as shown in Fig. 2, metaknowledge used by the meta-reasoner may be directly distributed over the domain knowledge structures rather than being confined to the meta-level. That is, the EVPs are encoded as part of an agent's AN. As we described in (Jones & Goel, 2012), the Augur system implements and evaluates the above architecture, knowledge representation, and computational technique in multiple domains.[1] In (Jones & Goel, 2012), a comparison of Augur's AN-based error correction technique with similar techniques (e.g. ANN backpropagation) is presented, and the compatibility of ANs with various kinds of classification techniques (operating within the AN nodes) is also discussed.

However, our past work left several theoretical and experimental questions unanswered. What is the number of hypotheses expressible in an AN for a compositional classifier? Is the diagnostic procedure used by the meta-reasoner optimal? Under what conditions does the learning converge? Might

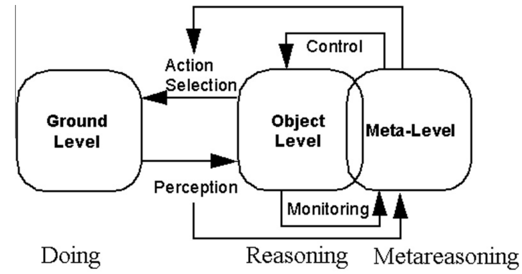


**Fig. 1** Our agent architecture for grounding meta-reasoning in perception.

[1] Although we use the term ''network'' in abstraction networks for generality, all of our work so far has focused on abstraction trees.