Case-Based Introspection*

Michael T. Cox
College of Computing
Georgia Institute of Technology
Atlanta, GA 30332-0280
cox@cc.gatech.edu

To effectively reason about one's own knowledge, goals, and reasoning requires an ability to explicitly introspect. A computational model of introspection is a second-order theory that contains a formal language for representing first-order processes and that processes instances of this representation. The reasoning algorithm used to perform such processing is similar to the algorithm used to reason about events and processes represented in the original domain: case-based reasoning.

Case-based understanding 1) takes as input some event in its domain along with its context, 2) based on salient cues in the input, retrieves a prior case to interpret the input, then 3) adapts the old solution to fit the current situation, and finally 4) outputs the result as its understanding of the domain. Similarly, case-based introspection 1′) takes as input a representation of some prior reasoning [e.g., an instance of case-based understanding] 2′) based on salient cues in the input, retrieves a prior case of reflection to interpret the input, then 3′) adapts the old case to fit the current situation, and finally 4′) outputs the result as its self-understanding. Here, the system's domain is itself.

We have extended the notion of an explanation pattern (XP) from Schank (1986) and Ram (1991). A *meta-explanation pattern* (Meta-XP) is an explanation of how and why an explanation goes awry in a reasoning system. We have developed two classes of Meta-XPs that facilitate a system's ability to reason about itself and to assist in selecting a learning algorithm or strategy. A *Trace Meta-XP* (TMXP) explains how a system generates an explanation about the world or itself, and an *Introspective Meta-XP* (IMXP) explains why the reasoning captured in a TMXP fails. The TMXP records the structure of reasoning tasks and the reasons for decisions taken in processing in a chain of decide-compute nodes. The IMXP is a causal structure composed of primitive, network structures that represent various failure types from a failure taxonomy. They are retrieved and applied to instances of reasoning captured in TMXPs and guide learning-goal formation after failure occurs.

Case-based introspection has proved useful during blame-assignment in a multistrategy learner called Meta-AQUA. Failure analysis cannot always look to the external world for causes. Often the assignment of blame is with the knowledge and reasoning of the system itself. Therefore, when Meta-AQUA encounters a reasoning failure while reading drug-smuggling stories, it uses case-based introspection to explain why it failed at its reasoning task. The system uses this analysis as a basis to form learning goals and subsequently to construct a learning plan to repair its memory. Figure 1 specifies the algorithm in some detail.

```
0. Perform and Record Reasoning in TMXP
1. Failure Detection on Reasoning Trace
2. If Failure Then
      Learn from Mistake:
        • Blame Assignment
            Compute index as characterization of failure
            Retrieve Introspective Meta-XP
            Apply IMXP to trace of reasoning in TMXP
            If Successful XP-Application then
              Check XP-ASSERTED-NODES
              If one or more nodes not believed then
                Introspective questioning
                GOTO step 0
              Else GOTO step 0
        • Create Learning Goals
            Compute tentative goal priorities
        • Choose Learning Algorithm(s)
            Expand subgoals
            Build learning plan
            Compute data dependencies
            Order plans
        • Apply Learning Algorithm(s)
```

Figure 1: Introspective Multistrategy Learning Algorithm (from Ram et al. 1993)

References

Schank, R. C. 1986. *Explanation Patterns: Understanding Mechanically and Creatively.* Hillsdale, NJ: LEA.

Ram, A. 1991. A Theory of Questions and Question Asking. *The Journal of the Learning Sciences, 1*(3&4), 273-318.

Ram, A., Cox, M. T., & Narayanan, S. 1993. Goal-Driven Learning in Multistrategy Reasoning and Learning Systems. In A. Ram & D. Leake (eds.), *Goal-Driven Learning*, Cambridge, MA: MIT Press. Forthcoming.

*This research was done with the author's advisor, Ashwin Ram.