MetaMAc, or What Do I Do Now?

A strategic perspective on autonomy beyond anomalies and goals

Don Perlis
Department of Computer Science
University of Maryland
College Park, MD, USA
perlis@cs.umd.edu

Michael T. Cox
Institute for Advanced Computer Studies
University of Maryland
College Park, MD, USA
mcox@cs.umd.edu

***Abstract*—In recent years, there has been strong interest in both reasoning about goal-identification and selection and metacognitive handling of anomalous situations. These two concerns are usually framed in terms of making agents more autonomous and flexible in dynamic and complex domains. Here, we wish to argue that there is a natural unifying perspective that includes both concerns and that may point the way to a yet more powerful kind of autonomy.**

Keywords—high-level autonomy; rational anomaly handling; goal reasoning; anomaly handling; monitoring, control, and management of autonomous self-aware systems

I. INTRODUCTION

An agent often has routine activities in which it is forming and/or following plans in pursuit of existing goals. And there are also situations in which it has to stop and ask itself: what do I do now? One major example of the latter is that of anomaly-handling: something seems out of the ordinary, contrary to expectation, and might indicate the need to do a form of error-correction. This has been the focus of much recent work, for instance Meta-AQUA [6], the Metacognitive Loop [2] and other similar efforts. Another example is goal-driven autonomy, in which an agent may autonomously alter or add to its goals if circumstances so warrant [1,8].

We wish to call attention to a level of processing at which an agent considers quite generally what to do: select from among several existing goals, form a plan to achieve an existing goal, continue with a current plan-in-action, alter such a plan, identify a new goal, abandon a plan or a goal, adopt new subgoals in response to unexpected events, explore opportunities for possible goals or other benefits, do a reality-check of beliefs and expectations, learn for learning's sake, and so on. This could perhaps be called the executive level of processing (borrowing that phrase from cognitive psychology), although that terminology already is in use in various cognitive architectures and so might not be the best choice. Instead, let us call this *reasoning at the strategic level*.

In what follows, we sketch some concepts related to the idea of such a processing level, argue that it usefully generalizes more traditional goal-reasoning and anomaly-handling, and outline what might be fruitful approaches to the strategic level. Section II simply states our hypothesis, and gives a key example; sections III –V describe existing work on rational anomaly-handling, goal-driven autonomy, and how they interact; VI discusses temporal issues, and VII presents conclusions.

II. METAMAC

We postulate a *metacognitive monitoring activity (MetaMAc)* that runs in parallel with an agent's normal routine activity of planning-acting in pursuit of already-identified goals. MetaMAc will be aware of such routine activities that are underway, and also of their aims and expectations, and of how (at least some) events are actually unfolding (which may or may not be as expected). As such MetaMAc represents the deliberate, conscious "self as process" monitoring and considering itself [3]. As MetaMAc processes this real-time information, it also asks itself over and over: What should I do now? What choices are there? Is there anything that would be better to do instead of (or in addition to) what I am doing? MetaMAc would normally run in the background, unless something pops into prominence in virtue of a certain salience or threshold that is reached.

We hypothesize that a MetaMac-enabled system would reveal considerable advantages over the same system without that enhancement. This could show up in many ways, but most especially in fewer errors over the long run.

Here is an example that illustrates various aspects of our idea:

A meeting is occurring, but participants are finding it hard to hear one another. One automated participant, X, identifies the problem as background noise coming from the hallway. X gets up and closes the door. Why? Because the closed door effectively blocks the hallway noise.

 ISBN: 978-1-61208-331-5

How does X come to be able to do this? While it may seem trivial, there are in fact a number of specific capabilities involved here, that illustrate our thesis:

1. X can identify a new problem on its own. This in itself is non-trivial; typically a human presents a problem to an automated system.

2. X can reason about causal relationships; but where does the causal info come from? (see item 4 below)

3. X can form new goals (problem solution ideas) on its own; but often there are multiple relevant goals, so how are they distinguished so as to pick the "best" ones (whatever that may mean)? Notice this, aside from closing the door, X could move closer to the speaker, ask the speaker to raise the volume, tape the mouths of the passersby, etc. So cost/benefit analysis is important.

4. X can and does pick up new information on a "knowledge is power" basis, independent of a specific immediate need. That is how X learned (long ago) that sound does not travel well through a closed door. But this presents complications as well: how does X decide when and what and how much to learn about "things in general?" There is an endless supply of such things, and X could easily become permanently absorbed in learning one narrow theme, or learning tiny random bits of distinct themes. Some sort of overall principles are needed here, perhaps in part guided by a concern to identify causal links.

III. RATIONAL ANOMALY HANDLING

No matter how complete an agent's knowledge base is or how good the agent is engineered, eventually mistakes or anomalies occur. In the face of surprise, an agent should be able to manage to adapt and do something reasonable. We call this capability *rational anomaly-handling (RAH)*. RAH is characterized by the following features [10]:

(i) We have expectations as to how things will be.
(ii) We compare expectation to observation and thereby note indications that an expectation has been violated.
(iii) We assess what we know that might explain this violation.
(iv) We decide what response, if any, to guide into place.
(v) We revise/create expectations as needed.

For the most part, artificial intelligence (AI) seems to be missing this key ingredient. After many decades of very fruitful work in machine learning, automated reasoning, planning, vision, natural language, and so on, we still do not have systems that come anywhere close to human-level performance of a general sort.

IV. GOAL-DRIVEN AUTONOMY

The idea of a goal tends to be conceptualized in two quite distinct ways in AI: as an end-state to be achieved, and as a kind of action to perform. In ordinary language, we conflate these, as in "I want to go to the beach," when we (presumably) mean "I want to be at the beach," which is a state-goal. Yet we also have maintenance goals, such as "keeping the room picked up" which presumably means vigilantly acting on any upcoming needs to do a picking-up action. In addition, there are intentions that are goal-like but perhaps not best described as goals. For instance, we can seek knowledge, in the belief that knowledge is a good thing to have; but no particular piece of knowledge can be identified as *the* goal here. Another is that of identifying what to do, if not particular goal is at hand; here the goal *could* be characterized as *find a goal*, but that begs the question in a way. Possible goals abound, all the time, so selecting one – if not at random – is a highly unspecified task, and could perhaps be driven by some agent-oriented measure of utility or interest.

The model we advocate - called *goal-driven autonomy (GDA)* [4] - casts agents as independent actors that can recognize problems on their own and act accordingly. Furthermore in this goal-reasoning model, goals are dynamic and malleable and as such arise in three cases: (1) goals can be subject to transformation and abandonment (2) they can arise from subgoaling on unsatisfied preconditions or in response to impasses during problem-solving and planning; and (3) they can be generated from scratch during interpretation.

For our purposes here, the most important of the above three cases is the third one. The idea is that given a problem in the world, an autonomous cognitive system must distinguish between perturbations that require a change in plans for the old goal and those that require a new goal altogether. What is missing in the planning and agent communities is a recognition that autonomy is not just planning, acting and perceiving. It also must incorporate a first-class reasoning mechanism that interprets and comprehends the world as plans are executed. It is this comprehension process that not only perceives actions and events in the world, but can recognize threats to current plans, goals, and intentions. We claim that a balanced integration between planning and comprehension leads to agents that are more sensitive to surprise in the environment and more flexible in their responses.

V. RAH AND GDA

RAH and GDA are closely related. RAH can lead to the conclusion that goals need to be altered or invented, for instance to avoid a repeat of a past anomalous situation (such as something identified as having prevented a goal from being achieved). This would then lead to invocation of a GDA process. And GDA in its own right can lead to the uncovering

 ISBN: 978-1-61208-331-5

of anomalies, such as goals not being met (which can trigger the application of RAH).

One way to envision this is in terms of the *A-distance* [7] which assesses alterations in time-series data that exceed a given threshold. This is a crucial kind of hedging-factor. For any given set of expectations will almost certainly fail to be fully identical to observed events. Tiny variations are the norm, and one cannot possibly attend to all of them (nor would it make sense to do so if it were possible). Yet how can suitable thresholds be determined, when context means everything? In some contexts, a small variation in color or noise-level may be insignificant, and in others may flag major problems or opportunities.

We think that learning is a promising approach here: an agent can learn, for a given context in which it may be operating (or planning to operate in), which are the important things to attend to. This can be partly at the explicit symbolic level (e.g., a teacher can tell the agent some items to watch for and some to ignore) and partly subsymbolic (experience can provide ranges of "normalcy" that the agent trains into its routines). While A-distance was developed largely for the latter situation, with continuous real-valued data, recent work has shown that it also is effective for discrete symbolic data [5].

Now, we doubt A-distance alone will be enough to cover all the cases that we envision for MetaMAc. For instance, another agent might simply tell our agent that something is important. That is unlikely to cross an A-distance threshold, since conversations may go on all the time, with words flowing rapidly back and forth. It would presumably require a rather high-level reasoning process to understand language sufficiently well to distinguish in a general principled way between, say, "such matters are unimportant" and "don't ever assume that such matters are unimportant."

The strategic level of MetaMAc then will likely require fairly sophisticated knowledge representation and reasoning (KRR) techniques; requiring reasoning about natural-language processing and the world more generally will at times be essential. Reference [9] gives compelling examples (although it is not focused on high-level strategic reasoning); see also [2] where three KRR criteria are given that flexible system should satisfy.

Thus, we suggest that the strategic level intended for MetaMAc involves an organizer of cognition that integrates activities and seeks to improve their effectiveness. But it also can involve entirely new concerns, such as the fortuitous discovery of a new homeostatic equilibrium in which things "work well" in totally unsuspected ways. This latter would not be an existing goal, nor one arrived at as something to achieve, but rather stumbled on and then perhaps adopted as a maintenance goal. The field of developmental (aka epigenetic) robotics would seem to fall into this category.

VI. Time and Other Realities

All processing takes time; and this applies to MetaMac as well. An issue immediately arises: how can MetaMac keep up with real-world changes (in its associated agent and in the world more generally)? A truly autonomous or autonomic system will not have the luxury of a human-in-the-loop, or of other activity being interrupted so that repairs can be made at leisure [11]. However, a special-purpose real-time logic is available [2], that was designed for such situations. Indeed, one of our aims in this research is to combine active logic with the ideas of MetaMac.

Indeed, the MetaMac idea is in effect a combination of two approaches that we have been exploring and implementing in recent years, namely the RAH and GDA themes. See [4,10] for details.

VII. Conclusion

In this paper, we suggest that an underlying research issue exists of considerable potential for enhanced autonomy: *how to design an agent with an effective and general-purpose "what do I do now" capacity*. This capacity bears on many cognitive processes and seems crucial for high-level reasoning in complex ever-changing environments. Researchers have at times studied aspects of what we describe under the MetaMAc banner in terms of metacognition (e.g., RAH). Other researchers have examined some of these issues in terms of goal reasoning and goal-driven autonomy. It may be the case that we are all speaking of the same process.

Acknowledgment

This material is based upon work supported by ONR Grants # N00014-12-1-0430 and # N00014-12-1-0172 and by ARO Grant # W911NF-12-1-0471. We thank the anonymous reviews for their comments and suggestions.

References

[1] D. Aha, M. Klenk, H. Muñoz-Avila, A. Ram, and D. Shapiro (Eds.) "Goal-directed autonomy: Papers from the AAAI workshop." Menlo Park, CA: AAAI Press, 2010.

[2] M. Anderson and D. Perlis, "Logic, self-awareness and self-improvement." Journal of Logic and Computation 15, pp. 21–40, 2005.

[3] J. Brody, M. T. Cox, and D. Perlis, "The processual self as cognitive unifier," Proceedings of the Annual Meeting of the International Association for Computing and Philosophy. 2013.

 ISBN: 978-1-61208-331-5

[4] M. T. Cox, “Question-based problem recognition and goal-driven autonomy,” D. W. Aha, M. T. Cox, & H. Munoz-Avila (Eds.), Goal Reasoning: Papers from the ACS workshop Tech. Rep. No. CS-TR-5029. College Park, MD: University of Maryland, Department of Computer Science. pp. 10-25, 2013.

[5] M. T. Cox, T. Oates, M. Paisner, and D. Perlis, “Detecting change in diverse symbolic worlds,” L. Correia, L. P. Reis, L. M. Gomes, H. Guerra, & P. Cardoso (Eds.), Advances in Artificial Intelligence, 16th Portuguese Conference on Artificial Intelligence, University of the Azores, Portugal: CMATI, pp. 179-190, 2013.

[6] M. T. Cox, and A. Ram, “Introspective multistrategy learning: On the construction of learning strategies.” Artificial Intelligence, vol. 112, pp. 1-55, 1999.

[7] D. Kifer, S. Ben-David, and J. Gehrke, “Detecting change in data streams.” Proceedings of the Thirtieth Very Large Databases Conference, pp. 180-191, 2004.

[8] M. Klenk, M. Molineaux, and D. W. Aha, “Goal-driven autonomy for responding to unexpected events in strategy simulations.” Computational Intelligence, vol. 29, no. 2, pp. 187–206, 2013.

[9] H. Levesque, “On our best behaviour.” IJCAI Research Excellence Award Presentation at the 23rd International Joint Conference on Artificial Intelligence, IJCAI-13, (Beijing), 2013.

[10] D. Perlis, “BICA and beyond: How biology and anomalies together contribute to flexible cognition.” Intl J of Machine Consciousness, v2, n2, pp. 1-11, 2010.

[11] D.Perlis, J. Elgot-Drapkin, and M. Miller, “Stop the world – I want to think”. International Journal of Intelligent Systems, 6, pp. 443-456, 1991. Special issue on temporal reasoning.

 ISBN: 978-1-61208-331-5