Opportunistic memory and visual search

Timothy M. Converse and **Kristian J. Hammond**
Department of Computer Science
The University of Chicago
1100 East 58th Street
Chicago, IL 60637
converse@cs.uchicago.edu

Abstract

In earlier work, we proposed a memory model that would facilitate the detection of opportunities to satisfy suspended goals. In this *opportunistic memory* model, suspended goals are indexed under feature sets that are predictive of the presence of the opportunity, and which are likely to be encountered in the normal course of future activity. The functional benefit of such encoding depends crucially on the particular vocabulary of features used, the costs of their detection, and the overlap of features relevant to the pursuit of different goals. In this paper we investigate the feature vocabulary implied by recent work on visual search [Treisman, 1985, Tsotsos, 1990], and its use in indexing goals suspended due to the lack of a particular object.

Opportunism

In daily life, people pursue a large number of different goals, and display great flexibility in shifting attention from one goal to another. These shifts in focus are often in response to new information that could not have been predicted in advance. We all capitalize on chance encounters with people we have been intending to talk to, buy things because we happen to see them displayed in a store, realize that we can perform one errand while "on the way" to perform another, and so on.

There is good reason to believe that a better understanding of this sort of opportunistic behavior is important for designing artificial agents as well. *Any* agent that must pursue multiple goals in an unpredictable world faces a problem of focus — which goals should be acted on now? If there are too many goals to plan for completely, or if not all of them are available when action must begin, then there must be some heuristic selection of a subset of goals that should be actively pursued. On the other hand, the agent needs to be ready to reconsider these choices in the face of unanticipated opportunities.

The central problem for accounts of opportunistic behavior is how to reconcile the background status of a goal that is not under active consideration with the fact that an opportunity to satisfy it can be recognized at all. Recognition requires perceptual and inferential work. If recognizing an opportunity to satisfy a background goal assumes as much work as actively seeking to satisfy it does, then the functional benefit of focusing on a small set of goals is lost.

Opportunistic Memory

In earlier work [Hammond, 1989b, Hammond *et al.*, 1988] we proposed a memory architecture that supports remindings of background goals at appropriate times. The idea of *opportunistic memory* relies on the fact that normal work toward the satisfaction of an active goal demands perceptual and inferential effort. In contrast to accounts of opportunistic behavior where goals are imbued with autonomous processing power [Birnbaum and Collins, 1984], the opportunistic memory model relies on encoding-time work to appropriately index passive notations in the same memory used in execution, so that suspended goals can be cheaply reawakened in the course of execution. Once "awakened", a suspended goal can be more thoroughly considered to see if it is plausible to pursue now. The potential benefit is that given proper encoding-time work, opportunity recognition can exploit the effort already being expended on goals that are currently active.

In brief, the opportunistic memory "algorithm" is as follows:

- Goals that cannot be immediately satisfied or integrated into a currently active plan are considered blocked, and are *suspended* [Schank, 1982].
- Suspended goals are associated with features in memory that, if encountered in the world, would indicate that the goal may now be satisfiable.
- The same memory structures are used to parse the world and make routine execution-time decisions. When the features that a suspended goal has been

associated with are activated in the course of this, the goal is brought back under consideration.

The objective of this encoding is that the agent will have goals "resubmitted" to attention by memory at exactly those times when they can be acted on. This will be successful to the extent that

1. the agent can characterize the ways in which the world would have to change to make the goal satisfiable, and
2. the agent can associate the goal with a set of easily computed operational features that strongly predict that such a change has taken place.

We now treat the first of these briefly, and then spend the remainder of the paper on the second.

Reasons for goal blockage

The simplest way that a goal can become blocked is if a standard plan associated with it is presently unusable for some reason. The possibility of encountering such a situation is an inevitable consequence of a case-based planning framework [Hammond, 1989a], where the attempt is made to reuse plans in situations that may not be exactly the ones for which they were designed.

Among the things which can block the use of a standard plan are:

- Lack of time.

  Due to other pressing goals and activities, there may not be time at the moment to run the standard plan. For example, the goal to run an errand to purchase something may have to be suspended when running late for an appointment.
- Lack of tool or resource.

  The plan depends on something we don't possess at the moment. For example, our plan to drink wine with our picnic lunch might be blocked by lack of a corkscrew.
- Lack of knowledge.

  In this case there is no state in the world that is blocking our efforts, but we lack the knowledge to properly exploit it. For example, the plan to get downtown via the infrequent commuter train might be blocked by lack of knowledge of the schedule.
- Lack of proximity.

  Being in a particular location is a common sort of prerequisite for a host of mundane tasks. Remembering that your plants need watering does not help you when you are away from home.

The features that should cause a goal to become reactivated don't have to be limited to those produced by a characterization of what originally blocked its satisfaction — it might be reasonable to index a goal under features that indicate unusual preconditions of a standard plan, on the theory that it is worth checking the other preconditions at that point. But the realization that a blocking condition is no longer true should be a strong predictor that the goal is now satisfiable.

Some of the categories above are abstract, and the corresponding opportunities are probably difficult to recognize (for example "having time"). From here on we will focus on opportunity recognition that corresponds simply to visually recognizing a particular object.

Functional requirements on indexing features

Clearly even with a good characterization of what would constitute an opportunity, the benefit of the encoding depends strongly on the particular vocabulary of features and the costs of detecting them. If a suspended goal is to be indexed under a given set of features, then

1. The feature set should be cheap to compute relative to the cost of completely verifying the existence of the opportunity.
2. As much as possible, the features should be ones that are likely to be checked anyway in pursuing other courses of action. (This consideration trades off with the previous one.)
3. The ratio of false negatives (missed opportunities) to false positives (inappropriate activations of suspended goals) should be in line with the urgency of the goal and the cost of verification.

We now discuss these requirements in the context of recent experimental results on visual search.

Visual search

It is extremely difficult to design experiments in cognitive and perceptual psychology that balance the conflicting demands of experimental control and relevance to real-world activity. Visual search is an attractive paradigm in part because, although people rarely face tasks exactly like the experimental task, it seems plausible that the faculty being tested is a building block (and possibly a bottleneck) in larger-scale visual activity.

In visual search experiments, subjects are given the task of quickly detecting a target image amid a field of distractors, and response time is measured. Treisman's interpretation of such experiments is to suggest that there is a small set of attributes of the visual field which are computed preattentively and in parallel [Treisman, 1985]. When the search task involves simply identifying that such a "pop-out" property is present, response times tend to be independent of the number of distractors. When identification of the target requires verifying a more complex property, or a conjunction of pop-out properties, response times vary linearly with the number of distractors, suggesting that serial attention is required in the verification. A small number of properties pop out, among them color, brightness, closure, curvature and tilt.

Tsotsos [Tsotsos, 1990] has argued for the intrinsic difficulty of *unbounded* visual search, where success depends on satisfying a function that, rather than being a straightforward match to a target, may depend on arbitrary interrelations of portions of the test image. Examples are tasks such as identifying the "odd man out" of a field of distractors, where the properties that will be unique are not identified in advance. Tsotsos proves an abstraction of unbounded visual search to be NP-hard, and uses this result to motivate a particular attentional architecture.

One implication of these results is, of course, that it is very unlikely that we can instantly identify all the objects in a complex visual scene. Ballard [Ballard, 1990] suggests that in general, the problem of simultaneously relating many different models to many visual locations may be too hard, and argues for the separation of *location* and *identification* algorithms, where location algorithms attempt to find a known object in the visual field, and identification algorithms do object recognition in situations where the location is not the issue. (In this he makes a functional argument for a specialization that is believed to exist in biological systems [Maunsell and Newsome, 1987, Mishkin, 1982, Mishkin *et al.*, 1983].) When incorporated in task-dependent behaviors, location and identification algorithms can use methods that are tailored to the outstanding attributes that are likely to be present in the context of the task, further avoiding visual reconstruction in its full generality. See *e.g.* [Swain, 1990] for real-time algorithms for identification or location of objects by their colors.

Visual search and opportunism

One way in which visual search experiments idealize real-world search is that, regardless of how the object of search is specified, finding the object is the only goal. It seems likely that in human activity, while there may often be a particular object being sought or identified, there are a number of other things that it is worthwhile to notice if they are seen. Visual cues that indicate opportunities to satisfy suspended goals may be among them.

An example

To motivate the rest of the discussion, we will tell a story:

> When I was in college, my black leather jacket was stolen. For some weeks after that, whenever I saw anyone wearing anything black and shiny, it "caught my eye", and I checked to see if it was my jacket. This happened even though, to my knowledge, I wasn't thinking about the jacket beforehand.

Depending on the importance of the goal of recovering the jacket, this is a functional response. The urgency and the chances of seeing the jacket at any one time don't warrant actively searching for it all the time. Still, if by chance it is encountered, it might be worth having a word with the current wearer.

Choosing visual indices

How should the opportunity represented by seeing the jacket be indexed in memory?

Our aim here is to have as much as possible of the work needed for opportunity recognition be done in the normal course of activity, with our suspended goal remaining a passive annotation. At the risk of drastic oversimplification, we can use the results cited above to sketch a hierarchy of effort that might be expended in recognizing an opportunity associated with a particular object.

1. Recognition of a particular pop-out property in the visual scene. For example, simply detecting an object of a certain color. Also, it may be possible to preattentively detect the conjunctive presence of more than one pop-out property, without necessarily verifying that they are in the same location or due to the same object.
2. Verification of a conjunct of pop-out properties in the same location, or any other process that demands serial visual attention, whether overt or covert.
3. True object recognition, i.e. classifying a seen object as a jacket or a fork or a car.
4. Inference based on object recognition. Here we move into processing that may or may not be completely non-visual, but that has the recognition of a particular object as a prerequisite. Note that much of the work of actually evaluating an opportunity represented by seeing an object falls into this category, since even if the object seen is exactly what was being sought at encoding time, circumstances may have changed so that the goal is not satisfiable for some other reason.

The point of this hierarchy is that, while the processes occurring at a given level may be very complex (and poorly understood), it seems reasonable to assume that the higher-numbered levels are dependent on the lower ones, and work at all levels will have to be done at some point to detect, verify, and evaluate opportunities represented by seeing particular objects. If a description that falls at one of these levels is associated with a suspended goal, we assume that activation of those features will be enough to activate the goal, which in turn will lead to the direct suggestion of the rest of the work necessary in its verification. So the question in determining indexing features then becomes: what is the lowest level at which a given goal should be indexed?

As we said above, a good recommendation for a potential index is that it will be computed anyway. Almost by definition, pop-out properties qualify (although detecting a particular one may well require some non-spatial attention, e.g. to a particular feature map). There are two reasons why pop-out properties may not provide good indices in particular cases. The first is simply that many of these properties are quite low-level and even in conjunction are not expressive enough to characterize interesting opportunities. The most promising ones seem to be motion, brightness, and especially color - but if the relevant type of opportunity is seeing a restaurant, then they do not suffice. The second reason is the possibility of an unacceptably high rate of false alarms.

In our example story, the suspended goal of recovering the jacket might simply have been indexed in association with the color black.[1] Whenever that color appeared in the visual field, the goal would be reactivated, and further verification that in fact the color was not due to the missing jacket would have to be performed, probably resulting at least in redirection of visual attention. One way to interpret the story is that in fact the goal was associated with the conjunction of a color and a texture, and that even that requirement gave a sufficiently high false alarm rate that it caused the experience to be memorable.

Indexing an opportunity at the level of full object recognition may offer a better characterization and result in fewer false alarms, but it may also then restrict discovery to things that are fully attended to and identified. This may be acceptable when the cost of completely missing the opportunity is low. It may also be acceptable when enough is known about the structure of future activities and tasks to predict that the given object *will* be attended to at some point [Hammond, 1990]. For example, it may be reasonable to index the goal of responding to a particular letter lost on a messy desk by actually seeing and identifying the letter, if you know that you will be cleaning the desk at some point in the future.

No strong theory of how to determine the appropriate level of indexing is offered here - in part it may just have to be a process of adding features to conjuncts or moving up the hierarchy given above in response to a high false alarm rate. (For a decision-theoretic approach to tuning opportunity recognition, see [Brand, 1991].) The one thing that seems clear is that if the feature set under which a goal is indexed needs to be refined, there is a functional benefit to starting on the overly inclusive side. It is easier to notice inappropriate remindings than it is to notice the lack of appropriate ones.

More realistic theories of attention and incremental object recognition doubtless allow for better-defined intermediate "hooks" on which annotations about opportunities can be hung. But to a great extent the lower-level visual vocabulary seems likely to be fixed, and choices of appropriate indexing level need to be made from the options provided by it; the visual system cannot be dynamically reconfigured to detect leather jackets preattentively.

[1] Here we make the simplifying assumption that the properties are binary—of course, the actual responses of "detectors" for the features are likely to be complicated and graded. The point here is that the goal is associated with some visual characterization of the object that can be detected preattentively.

Discussion

Relevance to artificial systems

All the discussion so far has been grounded in the particular feature vocabularies resulting from experiments on human vision. How do these ideas transfer to constructing systems that may operate under different perceptual constraints?

There seem to be strong design advantages to giving autonomous agents highly optimized peripheral perceptual systems, which compute a wide (but fixed) variety of low-level features all the time, very efficiently. More recently, there has been growing interest in the vision community in task-directed short-term control of perception, which actively directs the orientation of the perceptors, and uses attention to select a very small task-relevant subset of the provided information. At the same time, there is increased interest in the planning-and-action community in the realities and costs of perception

At its most abstract, our argument here is for creating indices for suspended goals on the fly, where the indices are composed of primitive perceptual features (whatever they happen to be) or features that may be derived due to attention to other goals. These indexed goals become activated when all their features are seen, but until then do not compete for attentional resources.

A closing story

There are types of opportunistic behavior which are not covered in the framework we've been discussing, one of which is exemplified by the following story:

> While on vacation at a lakeside cabin, the frisbee I was throwing became lodged in a tree. I moved down to the nearby shoreline to look for a small rock or piece of wood I could throw at the frisbee to dislodge it. As I looked, I saw a long paddle from a paddle-board, which I realized I could use to poke the frisbee free.

In effect, the opportunism here is in the recognition that the means to pursue an alternate plan is at hand. In part this is outside the framework we've presented because our focus has been on what goals should be attended to at all, rather than on suggestion of plans. The type of recognition that needs to be done is also subtly different, however. The recognition depends not only on object identification (as a paddle), but on further inference about the uses to which the object could be put (its suitability as a reach-extender), and as such falls into the highest category in the hierarchy we sketched above.

The only connection between the object and the current goal is through the plan that is eventually suggested. For some suggestions how of this integration of top-down and bottom-up inference might be implemented, see [Hammond *et al.*, 1990].

Acknowledgements

We'd like to thank Jim Firby, Patricia Goldweic, Mike Swain, Ron McClamrock, and Steve Whitehead for comments and stimulating conversations; they are not responsible for our errors and oversimplifications. We also thank anonymous reviewers for their comments.

References

[Ballard, 1990] Dana Ballard. Animate vision. *Artificial Intelligence*, 48(1), 1990.

[Birnbaum and Collins, 1984] L. Birnbaum and G. Collins. Opportunistic planning and Freudian slips. In *Proceedings of the Sixth Annual Conference of the Cognitive Science Society*, Boulder, CO, 1984.

[Brand, 1991] Matthew Brand. Decision-theoretic learning in an action system. Paper submitted to the ML91 Workshop on Learning Reaction Strategies, 1991.

[Hammond *et al.*, 1988] Kristian J. Hammond, Tim Converse, and Mitchell Marks. Learning from opportunities: Storing and reusing execution-time optimizations. In *The Proceedings of the Seventh National Conference on Artificial Intelligence*, pages 536–40. AAAI, 1988.

[Hammond *et al.*, 1990] Kristian Hammond, Timothy Converse, and Charles Martin. Integrating planning and acting in a case-based framework. In *The Proceedings of the 1990 National Conference of Artificial Intelligence*, August 1990.

[Hammond, 1989a] Kristian Hammond. *Case-Based Planning: Viewing Planning as a Memory Task*, volume 1 of *Perspectives in Artificial Intelligence*. Academic Press, San Diego, CA, 1989.

[Hammond, 1989b] Kristian Hammond. Opportunistic memory. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*. IJCAI, 1989.

[Hammond, 1990] Kris Hammond. Learning and enforcement: Stabilizing environments to facilitate activity. In *The Proceedings of the Seventh International Conference on Machine Learning*, July 1990.

[Maunsell and Newsome, 1987] J.H.R. Maunsell and W.T. Newsome. Visual processing in monkey extrastriate cortex. *Annual Review of Neuroscience*, 10, 1987.

[Mishkin *et al.*, 1983] M. Mishkin, L.G. Ungerleider, and K.A. Macko. Object vision and spatial vision: two cortical pathways. *Trends Neurosci.*, 6, 1983.

[Mishkin, 1982] M. Mishkin. A memory system in the monkey. *Philos. Trans. Royal Society London*, 298, 1982.

[Schank, 1982] R.C. Schank. *Dynamic Memory: A Theory of Reminding and Learning in Computers and People*. Cambridge University Press, 1982.

[Swain, 1990] Michael J. Swain. *Color Indexing*. PhD thesis, University of Rochester, November 1990. Technical Report 360.

[Treisman, 1985] Anne Treisman. Preattentive processing in vision. *Computer Vision, Graphics, and Image Processing*, 31, 1985.

[Tsotsos, 1990] John K. Tsotsos. Analyzing vision at the complexity level. *Behavioral and Brain Sciences*, September 1990.