A New Rational Framework for Modelling Exploratory Device Learning … but does it Fit with ACT-R?

Richard M Young & Anna L Cox

Psychology Department, University of Hertfordshire, Hatfield, Herts AL10 9AB, UK
{r.m.young,a.cox}@herts.ac.uk

Summary

We are investigating a new, rational framework for exploratory learning. We are applying it to model the exploratory learning of simple interactive devices — such as boxes with buttons and lights, or domestic central heating timers (Cox & Young, 2000) — although we would like to think that our framework is of considerably greater generality. (To some extent, exploration lies in the eye of the beholder, and it is possible to view many everyday cognitive activities as forms of exploration.) The framework promises to offer an account of various empirical phenomena in exploratory device learning, but — despite its rational basis — it is difficult to fit it onto ACT-R.

The framework

The rather scattered empirical literature on exploratory learning contains hints that people behave somewhat differently when performing *free* exploration of a device (e.g. when told to "try the device and see if you can figure out how to work it") as against *focused* exploration (i.e. when given a particular goal to achieve by exploration, such as setting the heating timer to switch on at a specified pattern of times). In focused exploration, people seem to learn more about the specific task but less about the device in general, and *vice versa*. For example, Trudel & Payne (1995, 1996) show that providing Ss with a "goal list" — i.e. an ordered list of tasks to try — leads to improved performance on tests following exploratory learning of a simulated digital watch. Rieman (1996) reports that people generally prefer to learn devices in the context of a particular task, a finding which begs to be tied to the common observation that, left to themselves, people typically learn only a small part of an interactive software system and fail to perform tasks in the most "efficient" way possible (e.g. Bhavnani & John, in press).

To understand our framework, it is best to imagine an agent engaged in free exploration. (The extension to focused exploration will be dealt with below.) We suppose that the agent has available, at any moment, a repertoire of possible exploratory acts that can be taken, which we designate as EA. Each EA may involve external action (such as pressing a button) or may be purely internal (as when pondering the significance of the label on a button, or of a particular lamp illuminating). Furthermore, we suppose that each EA may be "quantised", with just one quantum being performed at a time. In other words, when pondering the button label, the first quantum may simply notice that the label is the name of a colour and written in that colour, e.g. the label RED written in red. A subsequent quantum may notice that the label's colour corresponds to the colour of the lamp adjacent to it, and update various hypotheses and propose experiments in consequence.

In common with most other theories of rational behaviour, we assume a set of costs and values. Specifically, we assume that each quantum of an EA is characterised by

- its cost, C, which we usually identify with the time taken to perform it;
- its value, which is the increase of information, ΔI, estimated to follow from its execution.

(As is the way with such things, we do not specify further the nature of the ΔI.)

It should be noted that the value of a particular quantum of EA depends on what is already known. For example, when nothing is known about the function of a particular button, say, pressing it is likely to yield some valuable information. However, once the button is well understood, the estimated information to be gained from pressing it yet again falls to near zero.

It is known that the rationally optimum strategy in such circumstances is to pick at each stage the quantum of EA with the greatest *efficiency*, defined as the ratio $\Delta I/C$. (In return for making some further technical assumptions, a mathematical proof of its optimality is available [Ahlswede & Wegener, 1987, following earlier work by Kadane & Simon, 1977; Anderson, personal communication], but the strategy can also be understood in a commonsense way by imagining a series of such choices, and plotting the cumulative gain of information against the cumulative cost. The strategy just described yields a curve that climbs as fast as possible, thereby yielding the greatest possible amount of information for any given level of cumulative cost.)

Our rational framework for exploration therefore takes the form of an iterative cycle, in which

1. All applicable quanta of exploratory acts (EA) are found;
2. the one with the highest efficiency ($\Delta I/C$) is chosen;
3. that quantum is executed;
4. the cycle repeats.

Some exploratory behaviours, of course, are extended in time, consisting of a sequence of more basic exploratory acts. An "experiment", for example, may consist of pressing a number of buttons in sequence and then observing and interpreting the result. Such behaviours can be generated from within our framework by having each EA in the sequence propose the next one as part of its execution. To take what is probably the simplest example, the exploratory behaviour of "pressing a button to see what happens" consists of at least two EA quanta in sequence: first, to press the button, the execution of which also proposes … the second, which is to interpret any perceptible change in the device.

In this way, once the agent starts a multi-act sequence, we would normally expect it to continue to completion. However, no guarantee of such completion is given by the framework. Instead, at each step, the rationally optimal next EA is chosen, which may or may not be the next one in the sequence. So, on the one hand, a button press would normally be followed by an interpretation of its outcome. On the other hand, unexpected events (e.g. the button burns the agent's finger) will cause exploration to diverge from the sequence. The control structure is thus capable of executing these compound acts, but remains flexible, open, interruptible, and without commitment to higher-level control regimes. (Very much like a production system, in fact.)

Consequences of the framework

Sometimes this rational framework for exploration seems to us so simple and obvious that we worry whether there is any content to it. We reassure ourselves, however, by reminding ourselves how this simple, uniform framework removes the complexities and inconsistencies of other approaches we are familiar with (including our own), and how it neatly ties together a number of different aspects of exploration.

Instrumental behaviour

We return first to the topic of *focused* (as against *free*) exploration. Recall that our framework was developed on the assumption that the agent is engaged in free exploration, with no specific device-relevant target to achieve. Consequently, the only reason to take an external action (such as pressing a button) is because, on the grounds outlined above, it is a rationally optimal exploratory move to make. In focused exploration, however, where the agent has a specific goal to achieve, she may choose to take an external action for a quite different reason: namely, that she believes it will bring her closer to achieving her goal. We refer to this as an *instrumental* action, where a button is pressed, say, not for its exploratory value but in order to achieve an effect, e.g. causing a particular lamp to come on.

The framework provides a simple and elegant way to incorporate instrumental action into the story. In addition to the value an action may have by virtue of the estimated gain of information it will yield (ΔI), it may also have value because it is believed to lead to the achievement of the goal. (Such value might be computed as PG–C.) Thus, in focused exploration, actions have a different pattern of values than they do in free exploration, values determined in part by (the agent's beliefs about) the causal structure of the device, rather than by the estimated benefits for gaining information.

This way of incorporating instrumental action into the exploratory framework has several consequences.

- First, and most basically, the invocation of two different value structures responding to two different kinds of objectives, illustrates both the conceptual distinction between exploratory and instrumental behaviour, and also how they can be combined in a coherent account of rational action.
- Second, because the two values are responding to two different objectives (achieving the specific goal *vs* finding out about the device), different weightings are possible for the two kinds of objectives, reflecting different possible trade-offs between them.
- Third, as this mention of trade-off makes clear, there is inherent conflict between the exploratory objective and the instrumental objective. In other words, in general, because of the different value structures, the agent will not choose the same action for focused exploration as she would for free exploration. Or to put it yet another way, people will (in general) do and learn different things during free exploration than they will during focused exploration. (That is an interesting — and so far as we know, valid — empirical conclusion to draw from the framework.)

We were confused, but now all is clear …

From the perspective of this new framework, our own previous approach looks pretty confused and unsatisfactory. Our IDXL model of exploratory menu choice (Rieman, Young & Howes, 1996) was implicitly (we now realise) predicated on the notion of instrumental action. The basic structure of IDXL was that one of the menu options had to be taken, and the job of the exploration was to figure out which was the right choice. That view may sound innocuous, but it has great difficulty in coping with the fact that choosing one of the menu options is itself an exploratory move that can yield useful information about whether it is the right one.

This new framework cuts through and renders irrelevant a number of "satisficing-like" arguments that occur in IDXL and similar approaches to exploration, including possibly some of the rational analysis underlying ACT-R itself. Those arguments typically take the form of worrying about when to cut short the evaluation of alternative courses of action because the cost of further evaluation outweighs the likely improvement in the decision eventually taken. For example, Young (1998) attempts to extend the familiar PG–C rational basis of problem solving to the case of exploratory search, but gets very tangled up with considerations of when it is no longer worth spending effort on evaluating the alternative options but one should instead simply take the most promising option. That analysis has difficulty dealing with the idea that an option may be chosen in order to gain information about how good it is.

In our new framework we slice the cake differently, and such difficulties do not arise. We simply repeatedly choose the quantum of exploratory action with the greatest efficiency, and execute it.

Effects of instructions

The framework provides us with at least an orientation for analysing the effects on exploration of different instructions (such as those for focused as against free exploration). As we have seen, the framework suggests an approach in which instructions are viewed as setting up a particular structure of values, which then influence behaviour through means of the rational mechanism. This can extend to quite subtle aspects of the exploration. For example, in the case of a lightbox with red, green, and blue lamps and their associated buttons, free exploration will not favour one colour more than another, and will tentatively generalise (and perhaps then test) anything it learns about one colour of lamp and button to the others. However, given the focused goal of getting, say, the red lamp to come on, information about the red lamp and button will be valued more highly than information about the other colours, so that particular kind of generalisation and testing is less likely to occur.

Existing models of exploration in ACT

The only closely related model in ACT that we know of is the chapter on scientific discovery in the ACT-R book (Schunn & Anderson, 1998). That model is of course not intended to do device exploration, but the task is at least loosely analogous. We are struck by the observation that the model employs a very rigid control structure. It uses a fixed cycle of steps, namely (a) to design an experiment, (b) to perform the experiment, and (c) to interpret the results of the experiment. Thus, as a model of exploration, it views experimentation as the only kind of exploratory move; or seen the other way round, it views all exploratory moves as experiments. We do not criticise the model for the task at which it is aimed, but its control structure is clearly too rigid to capture the variety and flexibility exhibited by people in other exploratory contexts.

Implementation and fit with ACT-R

We have started to implement models within the framework, in both ACT-R and Soar, for a number of different devices. For ACT-R, the obvious implementation tactic is to use ACT-R's own mechanism of rational choice to select the next exploratory act. However, our exploratory framework doesn't fit too well onto ACT-R's machinery:

- ACT-R's rational choice mechanism applies specifically to the selection of productions, not of subgoals or other representations of exploratory acts (EA). Because the parameters associated with a production change only slowly (with long-term learning) and therefore cannot reflect dynamically changing circumstances, it is therefore necessary to artificially introduce a number of different productions for each EA, each one hand-coded to correspond to a particular context in which the EA might be invoked. For EA of any complexity, this will lead to a large number of ad hoc productions.
- ACT-R's rational calculation is based on PG–C, which has a different arithmetical "shape" to the efficiency measure $\Delta I/C$. It is therefore difficult to approximate the efficiency calculation with the ACT-R mechanism. It is tempting to use logarithms so that the subtraction in PG–C corresponds to the division in $\Delta I/C$. One can certainly consider identifying PG with $\log(\Delta I)$, but the trouble then is that the C in PG–C has to be identified with log(C) in the efficiency measure, which is a shame because those Cs are more naturally interpreted as the same quantity in both cases.
- Further, it is difficult to see how ΔI (or $\log(\Delta I)$) can be represented sensibly by the production parameters. The quantity P is computed as q*r, the product of two probabilities-of-success. It is unclear how the notional quantity "gain of information" is to be interpreted as a probability — the gain would have to be interpreted as odds, or something like that. It all looks very contrived.
- It is also very unclear how one could tell a story about how either the productions themselves or their parameters get learned. With regard to learning the productions, ACT-R is notoriously weak at learning to "specialise" productions by adding extra conditions (although ACT-R -5 may prove better in this respect). With regard to the parameters, it is plausible that the cost C could be learned from experience, since with appropriate care over subgoaling, it does actually correspond to the cost of performing the EA subgoal. On the other hand, as we have just seen, the notion of "gain of information" does not seem to correspond at all to a relative frequency of success, which is what the production probability parameters reflect.
- One other problematic correspondence is worth mentioning, although it goes beyond aspects of the exploratory analysis mentioned so far. In exploratory search, it is often important to combine multiple items of evidence, each of which may be individually inconclusive, into relatively strong evidence about a particular option. Bayesian updating provides a convenient mechanism for updating the explorer's stength of belief in a particular proposition. Now, that suggests the idea of using ACT-R's mechanism for learning associative strengths, which (to the extent that we understand it) uses more or less pure Bayesian updating. However, it is far from obvious that the representations can be twisted around in such a way that beliefs correspond to associations.

Conclusions

This talk reports work very much still in progress. In the longer term, the proposed framework will have to be judged by the quality of the models built within it.

In the meantime, the framework seems to be doing a useful job:

- It provides us with a clear, simple, and *rational* control structure for our models of device exploration.
- It provides a straightforward *a priori* explanation for some of the central (and perhaps puzzling) empirical findings, such as that people learn different things from focused as against free exploration.
- It provides a way of thinking about exploration clearly, something which — for us at least — was previously lacking.

However, fitting the framework onto existing ACT-R mechanisms is problematic, and appears to stretch the architecture uncomfortably.

References

Ahlswede, R. & Wegener, I. (1987) *Search Problems*. Wiley.

Bhavnani, S. K. & John, B. E. (in press) The strategic use of complex computer systems. *ACM Transactions on Computer-Human Interaction*, in press.

Cox, A. L. & Young, R. M. (2000) Device-oriented and task-oriented exploratory learning of interactive devices. In N. Taatgen & J. Aasman (eds.), *Proceedings of the Third International Conference on Cognitive Modelling*, 70-77. Veenendaal, The Netherlands: Universal Press.

Kadane, J. B. & Simon, H. A. (1977) Optimal strategies for a class of constrained sequential problems. *Annals of Statistics, 5*, 237-255.

Rieman, J. (1996) A field study of exploratory learning strategies. *ACM Transactions on Computer-Human Interaction, 3*, 189-218.

Rieman, J., Young, R. M. & Howes, A. (1996) A dual-space model of iteratively deepening exploratory learning. *International Journal of Human-Computer Studies, 44*, 743-775.

Schunn, C. D. & Anderson, J. R. (1998) Scientific discovery. Chapter 11 of: J. R. Anderson & C. Lebiere, *The Atomic Components of Thought*, 385-427. Mahwah, NJ: Lawrence Erlbaum.

Trudel, C.-I. & Payne, S. J. (1995) Reflection and goal management in exploratory learning. *International Journal of Human-Computer Studies, 42*, 307-339.

Trudel, C.-I. & Payne, S. J. (1996) Self-monitoring during exploration of an interactive device. *International Journal of Human-Computer Studies, 45*, 723-747.

Young, R. M. (1998) Rational analysis of exploratory choice. In M. Oaksford & N. Chater (Eds), *Rational Models of Cognition*, 469-500, Oxford University Press.