Modeling Automaticity and Strategy Selection in Dynamic Visual Detection

Bradley J. Best
Carnegie Mellon University

Cleotilde Gonzalez
Carnegie Mellon University, conzalez@andrew.cmu.edu

Michael D. Young
University of Colorado at Boulder

Alice F. Healy
University of Colorado at Boulder

Lyle E. Bourne, Jr.
University of Colorado at Boulder

Follow this and additional works at: http://repository.cmu.edu/sds
Modeling Automaticity and Strategy Selection in Dynamic Visual Detection

Bradley J. Best  
Cleotilde Gonzalez  
Dynamic Decision-Making Laboratory  
Social and Decision Sciences  
Carnegie Mellon University  
5000 Forbes Ave.  
Pittsburgh, PA 15213  
412-268-6242  
bradjbest@gmail.com, coty@cmu.edu

Michael D. Young  
Alice F. Healy  
Lyle E. Bourne, Jr.  
University of Colorado  
Department of Psychology  
Boulder, CO 80309-0345  
303-492-5032  
mdyoung@psych.colorado.edu, ahealy@psych.colorado.edu, lyle.bourne@colorado.edu

Keywords: Automaticity, Strategy, ACT-R, Cognitive Model, Dynamic Decision Making

ABSTRACT: A model of human automaticity in a dynamic visual task (RADAR) is reported. This model represents a first approximation for a model that reproduces the main results from human participants collected by Young, Healy, Gonzalez, and Bourne (2007). The initial results on response time demonstrate high degree of fidelity ($R^2=0.98$) using a process model derived directly from task requirements. The model leverages the ACT-R cognitive architecture (Anderson & Lebiere, 1998) and depends primarily on assumptions that have been imported from previous modeling efforts. The model is an instantiation of a dual-process theory of automaticity that is mediated by a strategy selection stage.

1. Introduction

Several real-world tasks, ranging across fields such as aviation, military and healthcare, require participants to develop highly skilled and automated levels of performance to achieve fast and accurate responses to critical stimuli in the environment. In the context of target detection tasks, such automated detection processes are particularly relevant to situations such as that of a pilot detecting the presence of an enemy among “friendly” aircraft, a physician detecting the presence of a tumor in x-rays, or a luggage screener detecting the presence of a hidden weapon among objects in passenger luggage (Gonzalez, Thomas, & Madhavan, 2006).

The development of automaticity in such contexts is particularly important as these complex tasks are characterized by multiple target stimuli and distractors, and environmental variables such as time pressure and workload make these tasks extremely difficult to perform in the absence of a practiced skill set (Gonzalez et al., 2006; Gonzalez & Thomas, 2006).

The phenomenon of automaticity was largely defined by Schneider and Shiffrin (1977) and Shiffrin and Schneider (1977), using a visual search and detection paradigm. Human behavior was characterized using a dual-process theory where one process was labeled as controlled, and one was described as automatic. A controlled process can be characterized as slow,
deliberate, voluntary, serial, and requiring attention. Controlled processes are sensitive to the workload or number of items in the task. On the other hand, automatic processes can be characterized as fast, involuntary, parallel, and requiring little overt attention. Automatic processes, due to their parallel nature, can be insensitive to the number of items involved in a task (see Wolfe, 1998, for a review of these terms in the context of visual search).

Initial performance in a novel task is typically performed in a controlled fashion. However, given extended practice in a task where there is a consistent mapping between stimulus and response, where a target appears only as a target and never as a distractor, automaticity tends to emerge. The transition from controlled to automatic processing can be detected by a shift in the pattern of response times to process a set of items: When increases in the number of items do not produce proportional increases in response times, automatic processing is said to be occurring. This shift can also be described as a reduction in cognitive load. Automatic processing is not only faster; it is also less effortful due to lower attentional demands.

Shiffrin and Schneider (1977) and Schneider and Shiffrin (1977) identified mapping conditions that enabled the emergence of automaticity and labeled these conditions consistent mapping (CM), which means stimuli are consistently either targets or distractors and are not exchanged between sets, and varied mapping (VM), which means stimuli can appear as either targets or distractors.

In addition to depending on consistent mapping conditions, the emergence of automaticity can also be impacted by the addition of other tasks that compete for resources during the acquisition phase. Wickens (1980, 1984) demonstrated that competing tasks that share the same modality (e.g., two visual tasks) can be expected to produce more interference than tasks that depend on different modalities (e.g., one visual task, one aural task).

From a training perspective, the emergence of automaticity is often a desirable outcome: Performance becomes less effortful allowing both quicker performance and simultaneous completion of other tasks that might compete for attention or other resources that voluntary controlled task execution requires. However, automaticity might also have drawbacks in task environments where controlled processes might be called for due to, for example, a high cost for incorrect decision-making.

Although automaticity is implicated in many rapid decision processes, relatively few efforts have been made to understand the exact information used in making a rapid decision. One notable exception is the research program described by Schunn, Reder, Nhousvanisvong, Richards, and Stroffolino (1997). In a mental arithmetic paradigm, they investigated the decision of whether to calculate or retrieve from memory. In their study, participants were asked to make decisions about the method they would use to solve a math problem more quickly than they could actually make the decision itself. (They had to decide whether to retrieve the answer from memory or to calculate the answer upon very brief exposure to the problem.) One of their main findings was that the surface features of the items being processed drove the choice of which decision process to use, and participants could be tricked into attempting to retrieve an answer that they had not memorized if it shared enough surface features (e.g., operands) with a previously memorized problem.

Studying the interaction of these aspects of automaticity – the impact of dual tasking in different modalities, the contribution of surface features to decision making, the effect of environmental complexity, the effect of consistency of mapping – calls for a dynamic, complex task environment within which automaticity can be thoroughly evaluated. The RADAR task environment (Gonzalez & Thomas, 2006) provides just such an environment (described in detail below) in which the acquisition of automaticity can be carefully studied.

The research reported in this paper is part of a major effort to create computational cognitive models that can be used as predictive tools for the effects of empirically-based training principles. An associated research effort is that of ACT-R models of training data entry skills (Gonzalez, Fu, Healy, Kole, & Bourne, 2006).

The research reported in this paper leverages empirical research of the training difficulty hypothesis (Schneider, Healy, & Bourne, 2002). An experiment, which will be summarized below, was designed to test the hypothesis that training under more difficult conditions would enhance testing (Young, Healy, Gonzalez, & Bourne, 2007). The ACT-R model designed and reported here is a first approximation for a model that reproduces the main results found in such an experiment.
In what follows, we will introduce the RADAR task, explain the empirical results, explain the cognitive model, and present the results from both modeled and human participants. Finally, implications for training will be discussed. The purpose of this modeling effort is to create an explicit account of controlled and automatic processing on a dynamic visual task to improve our understanding of how to leverage these processes in a training context.

2. The RADAR Task

The RADAR task is a single-user control task in which the goal is to detect and eliminate a hostile enemy aircraft by selecting an appropriate weapon system (a screenshot of the RADAR task is shown in Figure 1). This task has been used to study the effects of automatic detection on decision making (Gonzalez & Thomas, 2006) and the effects of response mapping on automaticity development (Gonzalez et al., 2006).

The RADAR task has two components: (1) visual detection and (2) decision-making and an optional, tone counting component explained below. The research reported in this paper utilizes the visual detection component and the tone counting task, but a complete description of RADAR including the decision-making component can be found in Gonzalez and Thomas (2006).

The visual detection and memory component requires the user to memorize a set of targets and then look for the presence of a target on a RADAR grid. This component essentially reproduces the goals of Schneider and Shiffrin’s (1977) task, except that the visual elements in RADAR are not static but instead are dynamic. In Schneider and Shiffrin’s (1977) task, all stimuli appear within foveal vision, whereas in the RADAR task, the stimuli move on the screen, and thus eye movements and visual scanning are required to find a target. A target threat may or may not be present among a set of moving blips that represent incoming aircraft. The blips—in the form of symbols, digits, consonants, or blank masks—begin at the four corners of the RADAR grid and approach the center at a uniform rate. The detection of an enemy aircraft must occur before the blips collapse in the middle of the grid.

In addition to visual detection component, some participants are also asked to detect and count the number of auditory tones presented that deviate from
a given reference tone. Participants were individually calibrated to their own standard and deviant pitches prior to testing. These tones are presented irregularly at roughly 1 s intervals throughout the set of trials, so this count must be maintained and accumulated across the presentation of the set of individual frames within a trial.

A set of targets containing either 1 or 4 letters or digits (depending on condition), drawn from the 9 possible digits and a specific subset of 9 letters, is presented to participants prior to each set of trials (see Figure 2) and must be maintained in memory. The two main conditions of the experiment, consistent mapping and varied mapping, are defined by the keeping a member of the memory set consistently defined as target throughout a series of trials and not as a distractor (CM), or considering a memory element a target in some trials and distractor in other (VM). In the varied mapping (VM) condition, digits or letters could be targets (depending on the assigned target type, which was varied between subjects and always differed from that used for consistent mapping (CM)), whereas the same class (digits or letters) served as distractors. In the consistent mapping condition, either digits are targets and letters are distractors, or letters are targets and digits are distractors. Thus, if the targets are digits, a blip filled in with a digit is necessarily a target on a CM trial.

3. Human Performance on the RADAR Task

Young, Healy, Gonzalez, and Bourne (2007) collected an extensive data set recording human performance on the RADAR Task. The current modeling effort focuses on a subset of these collected data that spans two experimental sessions, with the second session conducted 1 week after the first. These sessions are subsequently described as “training” and “test”. Consistent mapping target type was varied between subjects, and was either digits or letters, whereas mapping condition (CM or VM) and cognitive load – either 1 or 4 blips filled in and either 1 or 4 targets – were varied within subjects.

Forty-eight subjects participated in this experiment and completed the two sessions. Each session took approximately 2.5 hrs. The blocks were presented in the following predetermined order: CM 1-1, CM 4-4, VM 1-1, VM 4-4, Break, VM 4-4, VM 1-1, CM 4-4, and CM 1-1, where the first number represents the memory set size and the second number is the frame size (the number of non-blank blips). These two parameters represent the workload in the RADAR task. 1-1 is the lowest workload, whereas 4-4 is the highest workload. Each block consisted of 20 shifts each, where a shift contained 9 individual frames. At the start of each shift a set of targets (either letters or digits) were presented for memorization, after which the frames started appearing. Both the first and last frames were blank, whereas the remaining frames contained either 1 or 4 filled in blips according to the experimental condition. Blips originated at the corners of the display and rapidly moved toward the center.

75% of the shifts had a target present, and each shift had 7 frames that had to be checked for targets. If no targets were presented, participants submitted a quiet airspace report by pressing the space bar.

4 conditions in the experiment varied whether subjects heard and counted deviant tones during training and during testing. The 4 conditions were: (a) tone during training and testing, (b) tone during training, silent during testing, (c) silent during training, tone during testing, and (d) silent during training and testing. There were 12 subjects in each of these 4 conditions. At the end of the shift, participants who heard tones then entered the number of deviant tones.

A measure of human performance on the RADAR task is the time required to identify a target on a target present frame. The following table presents one of the results from the experiment. The table shows the response time by condition for both consistent
mapping (CM) and varied mapping (VM) conditions, and whether workload was low (CM 1-1, VM 1-1) or high (CM 4-4, VM 4-4).

<table>
<thead>
<tr>
<th>Condition</th>
<th>RT at Training</th>
<th>RT at Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>CM 1-1</td>
<td>707.761</td>
<td>681.508</td>
</tr>
<tr>
<td>CM 4-4</td>
<td>1004.740</td>
<td>953.060</td>
</tr>
<tr>
<td>VM 1-1</td>
<td>683.304</td>
<td>717.141</td>
</tr>
<tr>
<td>VM 4-4</td>
<td>1242.503</td>
<td>1248.069</td>
</tr>
</tbody>
</table>

In general, the significant effects are of mapping condition (CM is faster than VM) and load (1-1 is faster than 4-4). There is a significant interaction between load and mapping condition – CM 4-4 is much faster than VM 4-4 whereas CM 1-1 and VM 1-1 are comparable. These results are consistent with the automaticity theory and with previous automaticity results using the RADAR task (Gonzalez & Thomas, 2006).

4. An ACT-R model of automaticity in the RADAR Task

One of the main challenges faced in modeling this task is to account for the observed pattern of results through an explicit mechanism. To achieve this goal, a detailed process model was constructed using ACT-R 5.0 (Anderson & Lebiere, 1998). Although many readers may be familiar with ACT-R, a brief overview will be provided for those unfamiliar with the architecture.

4.1 ACT-R

ACT-R (Anderson & Lebiere, 1998) is a unified theory of cognition developed with over 30 years of cumulative improvement. At a fine-grained scale it has accounted for hundreds of phenomena from the cognitive psychology and human factors literature. The version employed here, ACT-R 5.0, is a modular architecture composed of interacting modules for declarative memory, perceptual systems such as vision and audition modules, and motor systems such as a manual module, all synchronized through a central production system.

ACT-R is a hybrid system combining a tractable symbolic level (production system) that enables the specification of complex cognitive functions, with a subsymbolic level that tunes itself to the statistical structure of the environment. The combination of these aspects provides both the broad structure of cognitive processes and the graded characteristics of cognition such as adaptivity, robustness, and stochasticity.

The central part of the architecture is the production module. A production can match the contents of any combination of buffers, including the goal, which holds the current context and intentions, the retrieval buffer, which holds the most recent chunk retrieved from declarative memory, visual and auditory buffers, which hold the current sensory information, and the manual buffer, which holds the current state of the motor module. During the matching phase, production rules whose conditions match perfectly to the current state of various information buffers (goal, memory retrieval, perceptual, etc.) qualify to enter the conflict set. Because ACT-R specifies that only one production can fire at a time, the rule with the highest expected utility from among those that match is selected as the one to fire.

4.2 RADAR Process Model Structure

Figure 3 shows the process flow within the model of human performance on the RADAR task. The processes specified support the rehearsal of the memory set, allocation of attention to the presented frames, encoding and processing of the presented targets and distractors, and determination of whether a target is present or not on the given display.

The model stipulates different processes for CM and VM conditions. This model is, in effect, making a claim that participants are aware of the experimental condition in which they find themselves. This awareness can be achieved through recognizing errors (false alarms) when they occur (see discussion for more on this). The effect of this awareness is that the overall process model is pruned, and is specific to condition. Each of these pruned process models will now be described individually.

4.3 RADAR Process Structure Specific to Consistent Mapping Conditions

The essential aspect of the model behavior when it recognizes a CM condition is that it does not attempt any memory retrievals. If a candidate item is of the same type as the memory set, it is identified as a target. If more than one candidate is present (in the 4-4 conditions), each is attended in turn. Thus, timing predictions from this condition are contingent on shifts of attention, but not retrievals from memory.
4.4 RADAR Process Structure Specific to Varied Mapping Conditions

During VM conditions the processing is more complicated. All candidates are the same type as the target (e.g., the candidates are letters and the memory set consists of letters). Hence, for all candidates a memory retrieval must be performed to make sure the candidate is not a distractor. This procedure produces the essential differences of the model in the VM condition: memory retrieval repeated for the number of items considered. In the VM 4-4 condition, the search will complete immediately if a target is identified, but otherwise it must be completed for each candidate. Given random ordering, this serial search will call for, on average, 2.5 memory retrievals. This is in contrast to the CM 4-4 condition, which will call for an average of 2.5 attention shifts, without the need for a retrieval.

4.5 Timing Assumptions used by RADAR Model

The model does depend on a set of explicit timing assumptions for producing detailed data. Initial perception of the frame and selection of target is assumed to take 350 ms, whereas each individual attention shift is estimated at 185 ms (Anderson & Lebiere, 1998). Adding the cost for an attention shift to the standard ACT-R cycle time yields 205 ms for any production that involves an attention shift. Motor action preparation and execution is estimated at 350 ms (making a response). The ACT-R architecture provides the retrieval times for items from the memory set which is impacted by both base-level activation and rehearsal (activation receives a boost during rehearsal). The time of memory failure is also important: If an item is not in the memory set, this value is determined by the retrieval “timing out,” which is impacted by the retrieval threshold. These assumptions, taken together, produce the timing predictions of the model.

4.6 RADAR Model Fit to Data

Figure 3. The structure of the ACT-R model of the RADAR task. The gray boxes represent those extra processes needed in the VM condition that are not necessary in the CM condition.
The process model described above, combined with the described timing assumptions, produces a model fit that is extremely close to the human data ($R^2=0.98$) for the combined training and test sessions. Figure 4 presents the average latency for both model and human data at test.

5. Discussion

This paper introduces an ACT-R model of automaticity that reproduces latency human data in a dynamic visual detection task, the RADAR Task. The model reproduces a common result in the automaticity literature: that the effect of workload is larger under the VM than the CM condition.

The issue of categorization and automaticity development has largely been a controversial issue (Cheng, 1985). The hypothesis concerning categorization that attempted to provide an alternative to the explanation of the automatization hypothesis by Schneider and Shiffrin (1977) has been clearly shown to be insufficient to explain the evidence of automaticity development (Schneider & Shiffrin, 1985). On the other hand, some accounts of automaticity have provided evidence that automaticity itself requires focal attention (Logan, 1992).

The particular process model described here suggests that participants are aware of the experimental condition (CM or VM) they are in to select their strategy for completing the task. That is, they are gating the decision process, and choosing whether to rely on a fast automatic process or a slower conscious process. Similar initial decision processes have been demonstrated in a range of empirical studies (e.g., Schunn, et. al., 1997; Lovett, 1998; Lovett & Anderson, 1996).

In addition to the present effort, other modeling efforts that have attempted to account for automaticity results (e.g., Anderson & Lebiere, 1998) have similarly depended on either awareness of experimental condition, awareness of stimuli category, or both, to replicate the results of Shiffrin and Schneider (1977) and Schneider and Shiffrin (1977).

Although this may not be a conventional interpretation of automaticity, the task requirements make it clear that there must be some mechanism that chooses between retrieving and deciding. Given that participants demonstrate evidence of acquiring automaticity in blocks of consistent mapping that are interleaved with blocks of varied mapping, it is not unreasonable to expect that there is some awareness of this choice. The categories are distinct (letters and numbers) for the CM condition and targets and distractors pertain to the same category in the VM condition. The current modeling effort indicates that awareness of this block structure (or at least a transition to a period of CM from VM) may facilitate development of automaticity by encouraging reliance on the automatic process.

Adaptation to such a shifting structure is a hallmark of human behavior, and has been previously modeled using the ACT-R architecture in other domains such as air-traffic control (Best, Schunn, & Reder; 1998) and an isomorph of the Luchins’ Water Jug Task (Lovett, 1998).

There is still a challenge to validate this process against the data and even the challenge of identifying what that validating data would be. Our future efforts will attempt to identify evidence of shifting strategies within the experimental blocks in the RADAR task, and seek the same data in other similar tasks. What we expect to find is that, within blocks that are VM blocks, false alarms will be clustered in the early portions of the block, after which participants will start to retrieve (the strategy shift being driven by failures). Conversely, we would expect CM blocks to be characterized by a gradual drop in response time across the block as memory retrievals are skipped (i.e., the VM strategy might persist but the more efficient CM strategy will gradually be adopted).

This model may have significant implications for training, specifically relating to transfer. Training in a VM condition is known to prevent automaticity.
However, if this model is correct, training in a CM condition can be expected to lead to a substantial number of false alarms when first exposed to a VM environment. Thus, training ought to proceed from one to the other: initial CM training to build up automaticity followed by a brief period of VM training to allow for trainees to learn to make the determination of when to pursue an automatic decision process. This kind of training hypothesis has been tested empirically (Gonzalez & Madhavan, 2006).

The true value of this modeling effort may be that we have been forced to make the model explicit in terms of cognitive operations. The main shortcoming of the current model, which we plan to address, is the need to allow the model to determine the appropriate strategy based on a pattern of changing success and failure and a drive to reduce cost (i.e., false alarms driving the transition to more effortful retrievals, and efficiency seeking driving the transition to skipping retrievals). Another useful enhancement would be the addition of the ability to learn individual stimulus specific rules within the ACT-R model through exposure (e.g., Best, 2006). In addition, richer validating data should be identified. If the current model, augmented with adaptive strategy determination, accounts for a pattern of shifting hits, misses, false alarms, and correct rejections, it will be an even more compelling model.

6. Conclusions

The model of the RADAR task presented here emulates the response time data of human participants reported in Young et al. (2007) with a high degree of fidelity (R squared = 0.98) using a process model derived directly from task requirements and depending primarily on assumptions that have been imported from previous modeling efforts. The model supports the view that dual-process theories of automaticity should be augmented with a strategic decision process that mediates between reliance on a fast, automatic process, or a slower, deliberate process.

7. Acknowledgements

This research was supported by the Multidisciplinary University Research Initiative grant from the Army Research Office (W9112NF-05-1-0153).

8. References


Schneider, V. I., Healy, A. F., & Bourne, L. E., Jr. (2002). What is learned under difficult conditions is hard to forget: Contextual interference effects in foreign vocabulary acquisition, retention, and transfer. Journal of Memory and Language, 46, 419-440.


Author Biographies

Brad Best is a Research Scientist in the Department of Social and Decision Sciences, Dynamic Decision Making Laboratory at Carnegie Mellon University where he focuses on cognitive modeling of adaptive behavior in complex environments.

Coty Gonzalez is an Assistant Professor in the Department of Social and Decision Sciences and the Director of the Dynamic Decision Making Laboratory at Carnegie Mellon University.

Michael D. Young is a graduate student in the Department of Psychology at the University of Colorado.

Alice F. Healy is a Professor in the Department of Psychology and Institute of Cognitive Science at the University of Colorado.

Lyle E. Bourne, Jr. is a Professor Emeritus in the Department of Psychology at the University of Colorado.