Using an Ethologically-Inspired Model to Learn Apparent Temporal Causality for Planning in Synthetic Creatures

Robert Burke
Synthetic Characters Group, The Media Lab, MIT
20 Ames St. E15-320H
Cambridge, MA, USA, 01239
(1)617-253-9832
rob@media.mit.edu

Bruce Blumberg
Synthetic Characters Group, The Media Lab, MIT
20 Ames St. E15-311
Cambridge, MA, USA, 01239
(1)617-253-9832
bruce@media.mit.edu

ABSTRACT
Inspired by recent work in ethology and animal training, we integrate representations for time and rate into a behavior-based architecture for autonomous virtual creatures. The resulting computational model of affect and action selection allows creatures to discover and refine their understanding of apparent temporal causality relationships which may or may not involve self-action. The fundamental action selection choice that a creature must make in order to satisfy its internal needs is whether to explore, react or exploit. In this architecture, that choice is informed by an understanding of apparent temporal causality, the representation for which is integrated into the representation for action. The ability to accommodate changing ideas about causality allows the creature to exist in and adapt to a dynamic world. Not only is such a model suitable for computational systems, but its derivation from biological models suggests that it may also be useful for gaining a new perspective on learning in biological systems. The implementation of a complete character built using this architecture is able to reproduce a variety of conditioning phenomena, as well as learn in real-time using a training technique used with live animals.

Categories and Subject Descriptors
I.2.6 [**Artificial Intelligence**]: Learning – *concept learning, induction, knowledge acquisition, parameter learning.*

General Terms
Algorithms, measurement, performance, design, reliability, experimentation, human factors, theory, verification.

Keywords
Autonomous agents, apparent temporal causality, ethology, synthetic characters, virtual creatures, planning, reactive systems.

1. INTRODUCTION
In order to survive in a dynamic environment, many self-regulating systems – both biological and computational – make use of representations that model important aspects of the world. Two such representations fundamental for living systems are the passage of time, and the rate at which they experience relevant stimuli.

Early models of behavioral conditioning, such as the Rescorla-Wagner model, minimized the use of representation and speak simply of animals forming and strengthening associations between stimuli. While that associative model is successful at rendering explainable certain phenomena, there is a wide range of phenomena it is unable to model without substantial trouble, such as the ability to learn an expected latency of reinforcement. Recent studies by Gallistel and others have considered the possibility that models of time and rate are fundamental to conditioning phenomena. Gallistel and Gibbon propose two new models – Scalar Expectancy Theory (SET) and Rate Estimation Theory (RET) – that require an animal to represent the length of the interval between stimuli, and the rate of reinforcement associated with various stimuli. Using these models, the authors are able to account for a number of conditioning phenomena that can not be explained using the Rescorla-Wagner model [1], and they do so in a clear and elegant way.

Similarly, much of the early work in behavior-based artificial intelligence minimized the importance of representation [2]. The Synthetic Characters group at the MIT Media Lab designs cognitive architectures for autonomous and semi-autonomous creatures that inhabit graphical worlds. By using ethological models to inform our design, we seek to extend the work and philosophy formulated by Blumberg [3]. We recently built a layered brain architecture for behavior-based virtual creatures [4].

Our next goal was to re-implement much of that system's learning and action-selection mechanisms in a way that paid attention to the sort of details that Gallistel attends to in the SET and RET models. The resulting representations and mechanisms needed to operate in real-time with dozens of potential stimuli. We wished to maintain, and hopefully improve upon, the system's ability to model a dog training paradigm and other sorts of learning.

We have arrived at the representations and mechanisms described here. They are not simply a recreation of SET and RET. Instead, they represent a hybrid that integrates new components inspired by Gallistel and Gibbon's work into the Synthetic Characters cognitive architecture. A creature constructed using this new architecture can predict and plan for future events by discovering causality relationships in the world. The creature is motivated to learn by a desire to satisfy its drives, and explain the salient stimuli it perceives. Its representation of apparent temporal causality is tightly integrated with its fundamental representation for action

AAMAS'02, July 15-19, 2002, Bologna, Italy.

selection.

1.1 Paper Overview

In Section 2, we introduce the sources of inspiration behind this work: the philosophy of the Synthetic Characters group, the layered brain architecture upon which we built this system, and the ethological model. We then describe the new cognitive architecture in Section 3. An example produced with our implementation illustrates the benefits that virtual creatures obtain by learning about apparent temporal causality on the job. In Section 4, we discuss our results from both autonomous agents and ethology points of view. In Section 5, we present references to some related work from both fields.

2. Background

2.1 Philosophy: Synthetic Characters Group

The Synthetic Characters group has recently sought insight into the nature of intelligent behavior by building characters inspired by the capabilities of dogs. Dog training is applied operant conditioning, and a domain in which one sees many of the phenomena described in lab experiments, but in the context of a whole "behaving" creature. One form of training, known as "clicker training," involves the use of a handheld device called the "clicker" that makes a short, sharp clicking noise. This noise serves as a precise event marker for the creature. When repeatedly followed by a treat, the noise of the clicker becomes associated with a food reinforcer. Clicker training has been successfully used to train animals ranging from dogs to dolphins [5].

2.2 Architecture: c4

The Synthetic Characters' agent-based cognitive architecture for building virtual creatures, detailed in [4], is composed of many fairly simple components, each individually unintelligent, but capable through their interaction of producing complex cognitive behavior. The systems in a creature's brain are divided by function into three parts: a representation of the internal and external worlds ("represent the world"), an action selection mechanism ("decide what to do"), and a navigation and motor system ("figure out how to do it").

2.3 Ethology: Time, Rate and Conditioning

Gallistel and Gibbon detail two theories that account for a broad range of conditioning phenomena [6]. These theories depend on an animal's ability to learn temporal intervals between events, as well as rates of reinforcement. What is exciting about these models is that by assuming the existence of representations for time and rate, Gallistel and Gibbon are able to easily explain a wide range of disparate conditioning phenomena.

One main goal of behaviorism is the identification of basic learning processes that can be described in terms of stimuli and responses. The experimental paradigm that underlies the study of conditioning is one in which the subject is presented with various stimuli. The subject learns associations between the stimuli.

In Classical (Pavlovian) conditioning, a stimulus that previously did not elicit a response comes to elicit a response after it is paired for one or more trials with a stimulus that already elicits a response.In Operant (Instrumental) conditioning the consequences of a response increase or decrease the likelihood that the response will occur again. In one such procedure, the subject learns that performing a certain behavior in a context results in a reinforcer such as food [7]. Most contemporary associative theorists no longer assume that the association-forming processes in Classical and Operant conditioning are fundamentally different. Rather, they are thought to give rise to different associative structures via a single association-forming process.

2.3.1 Scalar Expectancy Theory: "When"?

Scalar Expectancy Theory, or SET, pertains to the onset of the conditioned response (CR) following a stimulus onset, revealing both "when" and "for how long" the CR should occur. According to SET, when an animal perceives a salient stimulus, the creature starts an internal timer that records the (subjective) interval between that stimulus and another salient stimulus, such as a reinforcer. Later, when the first stimulus is perceived again, the animal starts another timer, and decides when to respond by using the ratio of the elapsing interval to the remembered interval. When the ratio exceeds a threshold, $_1$, which is close to but generally less than 1, the subject responds. The results produced by SET correlate with some well-established facts about how subjects time the duration between two events:

• The CR (which suggests the expectation of the second event) is maximally likely at the reinforcement latency

• The distribution of CR onsets and offsets is scalar, and thus the temporal distribution of CR initiations and terminations is time scale invariant. In other words, when one signal seems to predict a future event, the approximate size of the window in which a subject expects that event to occur increases with the interval.

2.3.2 Rate Estimation Theory: "Whether"?

SET assumes that the animal has already determined whether or not a stimulus merits a response. In the Rate Estimation Theory model, this decision is based on an animal's growing certainty that a stimulus has a substantial effect on the rate of reinforcement. Gallistel and Gibbon provide a computational model for how animals determine the true rates of reinforcement for each stimulus. The principle of parsimony – essentially Occam's razor – is used to find the simplest unique solution to the problem of determining the rates of reinforcement. Mathematical details are found in the appendices of [6].

2.4 Summary of Our Goal

The previous Synthetic Characters cognitive architecture could be said to integrate an analysis of the past with an ability to react to the present. With the new architecture, we sought to include a representation of the future inspired by SET and RET. Thus a creature could be *informed* by salient stimuli perceived in the recent past, *reactive* to stimuli perceived in the present, and *able to plan* appropriately for the stimuli predicted to appear in future. This augmentation would further our ability to create robust creatures that can adapt to and learn from a dynamic environment.

3. Cognitive Architecture

By itself, a representation for apparent temporal causality won't improve the life of a virtual creature. We need to consider how a creature might *use* its knowledge of apparent causality to influence its action selection and help satisfy its internal needs.

We therefore begin this section with the notion that creatures have *internal needs* they seek to satisfy. These are represented by Autonomic Variables that we combine together into a multidimensional DriveVector (Section 3.1). We then discuss the fundamental *action selection* choice: whether to explore, exploit or react (Section 3.2). To help the action selection mechanism make this choice, we integrate a *representation of apparent temporal causality* into the mechanism. This representation, the "Predictor," lets a creature reason about causality relationships between stimuli, thereby providing an understanding of cause and effect that can accommodate changing ideas in a dynamic world (Section 3.3). Finally, we show how these Predictors let us model the *effects on a creature's emotional state*, thereby facilitating learning in other parts of the architecture (Section 3.4).

3.1 Creatures must satisfy Internal Needs

Our atomic component of internal representation is the Autonomic Variable. Autonomic Variables each produce a continuous scalar-valued quantity. Most Autonomic Variables have *drift points* – values which they drift toward in the absence of other input. Some of the creature's Autonomic Variables represent Drives such as *hunger*. In addition to its drift point, each Drive also has a *set point*, the value at which the drive is considered satisfied. The strength of the drive is proportional to the magnitude of the difference between the set point and its output value. Associated with each Drive is a scalar drive multiplier that allows the creature to compare the importance of various drives. Over the course of a creature's existence, these multipliers might change, so that the creature can favor different drives at different times. This mechanism can create periodic changes in the drives (for example, to produce a circadian rhythm) and induce drive-based developmental growth over a creature's lifespan. Take the output of all the Autonomic Variables that represent Drives, and concatenate their scalar output values into a vector, and we have the DriveVector – a summary of the creature's current drive state.

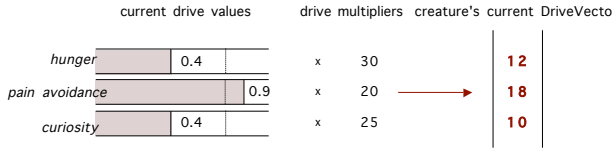

Figure 1: Three drives, multipliers, and resulting DriveVector.

Our creatures represent the value of something in the world – whether an action, a fellow creature, or an object – as a "value vector" with the same dimensions as the DriveVector. That vector indicates how the creature believes that thing will affect its drives. The utility of something in the world at a given moment can be reduced to a scalar value by taking the dot product of its value vector with the current DriveVector. The result parallels the motivational model described by Spier in that the utility of something reflects the creature's perceived drive state [18].

The creature described in Figure 2 has three drives: hunger, pain avoidance, and curiosity. These are concatenated into a three-dimensional DriveVector $|d_1\ d_2\ d_3|$. The creature's food source is a shed in which there are sleeping sheep. If he rattles the shed, the sheep will scatter and he can feast. However, the shed is surrounded by an electrified fence. Thus, in order to rattle the shed, the creature will have to endure a painful electric shock. The value of the "kick the shed" action (middle of Figure) might look like [-10 20 -3] relative to his drives [hunger, pain, curiosity], meaning that it will reduce his hunger drive (good), increase his pain (bad), and slightly lower his curiosity drive (because kicking stuff is intriguing). If this creature's current drives are [**5 5** 5] for [**hunger pain** curiosity], then the value of kicking the shed is [**5 5** 5] · [-10 20 -3] or 35, a *positive* number suggesting that, overall, the action will not be such a good thing, as it results in a net *increase* in drives. But, in the absence of other food sources, the creature's drives might eventually drift to [**10 4** 5]. Now hungrier and not in quite as much pain, the dot product of [**10 4** 5] and [-10 20 -3] generates a utility of -35; in other words, an effective strategy for satisfying the current drives. (We note that this example is functionally equivalent to a more mundane experiment wherein a rat is presented with a lever, surrounded by an electrified floor pad, which when pressed causes food pellets to be dispensed.)

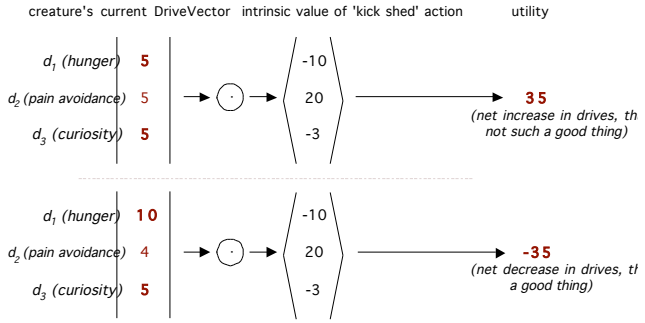

Figure 2: The utility of an action varies with the DriveVector.

Autonomic Variables are also used to model the creature's emotional state. We have worked with several models of affect that create a multidimensional "affective space." Each axis of the space is represented by an Autonomic Variable. Yoon describes the three-axis stance-valence-arousal model in [8] that was inspired by Russell [9] (see also Ekman's emotional states, [10]).

3.2 Action Selection: Explore, Exploit or React

The fundamental choice a creature must make at every moment is whether to ***exploit*** its knowledge about the world, ***explore*** the world to possibly discover new things, or ***react*** to recently-observed stimuli.

The action selection mechanism that will integrate these explore, exploit and react operations should exhibit the qualities suggested by Brooks in [11]. Every action performed by the creature should appear and be *relevant*. The creature's behavior should have a high degree of *persistence* and *coherence*, in that the creature should be aware of the appropriate duration of its actions and see them through to completion, without getting stuck in "mindless loops." The selection mechanism itself should be capable of *learning and adaptation*.

Figure 4 illustrates how the action selection mechanism integrates the Explore, Exploit and React operations. On every timestep, we first check if the creature needs to perform a reflex action ((1) in the diagram). If not, we check if the active action is completed (2). If so, the creature selects a drive on the basis of their relative magnitudes (3). If the curiosity drive is

chosen, the creature performs an Explore operation. If any other drive is chosen, the creature performs an Exploit operation, which is guaranteed to select a new desired action. After this, the React operation is performed on any newly-perceived salient stimuli, potentially causing the focus of attention and desired action to change (4). Thus a reaction can potentially interrupt the active action.

At the end of the timestep, the mechanism has in fact made three selections: it has chosen the desired action, the object of attention, and the target object. The desired action is a high-level token like "sit," "kick" or "approach" that describes what the creature should do. The target object is the object on which the desired action should be performed. The object of attention represents the creature's focus of attention. Each of these three selections is "winner take all," in that they are made to the exclusion of all other options for this timestep.

We now summarize the exploit, explore and reaction operations. Details and additional mathematics are found in [12].

3.2.1 Exploit

The Exploit operation causes the creature to use its knowledge about the world to select *consummatory actions* that it believes will help satisfy its drives, and *appetitive actions* that help move it closer to performing a consummatory action.

The creature can exploit by using its direct perceptions of the world to choose a new action state with a high utility. This is the typical action selection operation performed by a purely reactive autonomous agent.

A creature with an ability to predict future events can also exploit by using its TimeLine to react to something it predicts is about to happen. (The TimeLine, as shown in Figure 3, is a representation available to the entire action selection mechanism that organizes perceived salient events in the past and predicted events in the future.) If a painful stimulus is almost certainly about to appear, it should be avoided if at all possible. Similarly, if a stimulus about to appear will facilitate a consummatory action, the best course of action might be to approach the stimulus in preparation for its arrival. These represent *preventative* and *preparatory* actions.

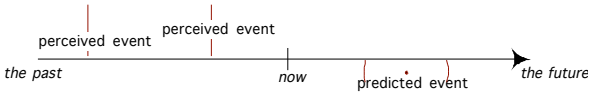

Figure 3: The TimeLine: Past, Present & Future Events.

The scalar utility values obtained for all options are used as input for a histogram probability distribution, from which the creature selects a single course of action in a winner-takes-all decision. The spirit of this mechanism is to typically cause the creature to select the very best available option, while still occasionally selecting another option that seems very promising but not necessarily the "best."

3.2.2 Explore

There are sufficiently many exploration techniques that, instead of peppering them throughout the action selection mechanism, we formalize our notion of exploration by encapsulating its many forms within the Explore operation. Potential avenues for Exploration include:

• Redirecting the creature's attention toward an interesting object.

• Exploring that interesting object by performing actions on it; perhaps randomly, or perhaps by selecting actions that produced useful results for similar objects.

• Testing predicting mechanisms in which the creature has low confidence, possibly by generalizing and discriminating the trigger contexts that cause them to make predictions.

• Selecting an unusual (rather than obviously useful) action state.

3.2.3 React

The react operation gives the creature a chance to interrupt its current behavior and react to the perception of a salient stimulus.

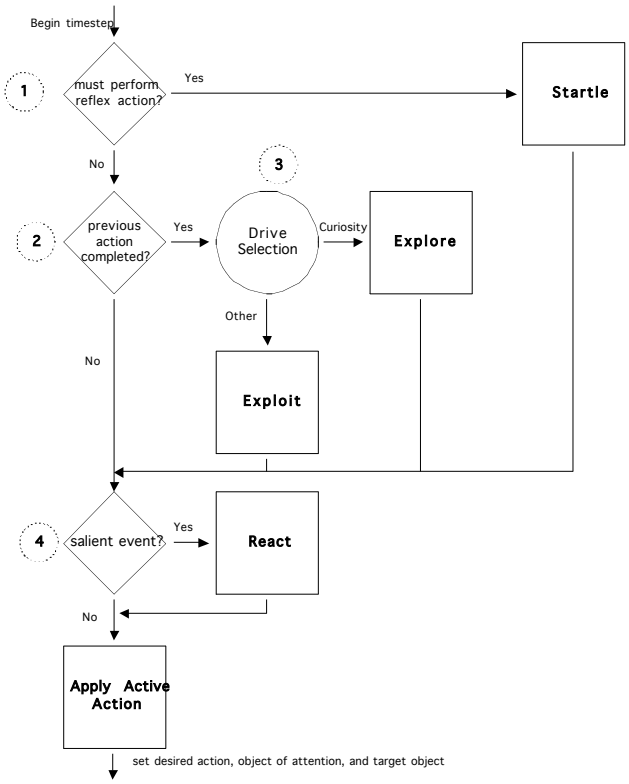

Figure 4: Action Selection Mechanism.

The first thing a creature does when it perceives a salient stimulus – for example, a loud noise – is *to try to explain it*, by looking at the TimeLine for a prediction of the event. The creature's response (both in terms of affect and action) will be largely determined by whether or not this event was predicted, and which action states, good or bad, the event is able to facilitate.

In addition to performing an action directly facilitated by the onset of a stimulus, the creature may choose to interrupt its current behavior in favor of one of essentially 3 suitable responses: to *approach, observe* or *avoid*.

If the creature can't explain the appearance of a stimulus, it is given an opportunity to invent an explanation, marking the beginnings of the apparent temporal causality process we will discuss in Section 3.3. For further discussion of "explaining away" unexpected events using probabilistic reasoning, see [13].

3.3 Apparent Temporal Causality

The Explore, Exploit and React operations assume that the creature has the ability to represent apparent temporal causality relationships. These are cause-and-effect relationships that a creature believes it has discovered in its world. They are *apparent*, because they are how the world appears to work to the creature, whether or not the world actually works that way. They are *temporal*, because, as in SET, cause and effect are somehow related in time. And they represent *causality*, in that the creature can use them to generalize from specific examples to arrive at general principles about how the world works. Similar temporal logic, as surveyed by de Kleer in [14], has been used in the past to extend the problem solving abilities of traditional planning systems [15]. As noted by Moray, four types of cause are classically distinguished (with classically meaning in the sense of going back at least to Aristotle). Here we are discussing an attempt to learn about *formal* causality, although extending this work to consider other causality types is an intriguing prospect [16].

3.3.1 First, represent Stimuli

A Stimulus is a signal provider in the creature's brain that can serve as a component of an apparent temporal causality relationship. The stimulus can thus represent a wide range of potential signals, from Percepts indicating external world state, to some component of self-action, to an Autonomic Variable representing a facet of the internal state.

Much discussion in behavioral psychology revolves around the animal's perception of the "onset" and "offset" of a stimulus, suggesting that at some point, the creature distinguishes between its presence or nonpresence. Thus every signal provider for a stimulus must provide an activation threshold.

3.3.2 Predictors represent causality relationships

Now that we have a means for representing stimuli, we need a way to represent causality relationships between those stimuli. A Predictor represents an apparent temporal causality relationship by recording the perceived interval between two events, where an event is defined as the onset or offset of a stimulus. The first event is recorded as the *Predictor Context*. The second event is recorded as the *Predicted Event.* The interval between the two events is recorded as the *Predicted Interval.*

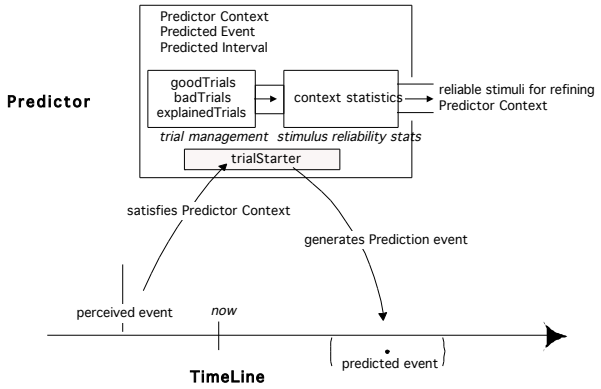

Figure 5: Predictor.

In Figure 5 we see the basic interaction between a Predictor and the TimeLine: when an event occurs that causes all of the stimuli that comprise the Predictor Context to become concurrently active, the Predictor begins a Trial, causing the expectation of a future event. Just as in SET, the size of the time window during which the event is predicted to occur is determined by a scalar function of the Predicted Interval. The importance of this property is depicted in Figure 6, where two Predictors with significantly differing Predicted Intervals produce predictions of events that are expected to occur within time windows of substantially different size. In that figure, the perceived event at the present time satisfies the Predictor Context for the two (unrelated) Predictors. Each begins a Trial, resulting in the prediction of two future events after their respective Predicted Intervals. The Predicted Interval for Predictor 2 is twice the length of the Predicted Interval for Predictor 1, and so the size of the windows in which the two future events are predicted reflects this.

Figure 6 also illustrates the mathematics of the decision threshold. $\__1$ somewhat less than 1 and $\__2$ somewhat greater than 1 are used to decide when the subject should respond. When the ratio ($t_e / i_{predictor}$) between the subjective duration of the currently elapsing interval (t_e, which has its zero at the time when the Trial begins) and the interval encoded in the Predictor ($i_{predictor}$) exceeds the decision threshold ($\__1$), the creature begins to expect the appearance of the predicted stimulus. When the ratio exceeds another threshold ($\__2$), the Predictor ceases to predict the event, and generates an expectation violation. Thus the Predictor effectively describes a "window" in which it predicts an event will occur. The window's dimensions are $\__1 i_{predictor} \leq t_e \leq \__2 i_{predictor}$, where $i_{predictor}$ is the Predictor Interval in the Predictor that began this Trial, $\__1$ is a creature-global constant slightly less than 1, $\__2$ is a creature-global constant slightly greater than 1, and t_e is the elapsed time since the Trial began (when the Predictor Context was met).

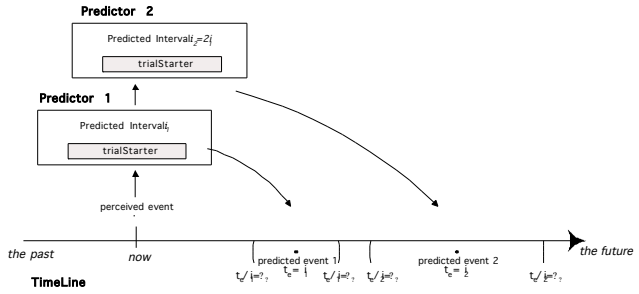

Figure 6: The timing of a Predictor window.

An ongoing Trial can expire in one of three ways. If the predicted event does occur during the time window as expected, the Trial is declared *successful*. If, without explanation, the predicted event fails to occur within the time window, the Trial can be declared a *failure*. If the predicted event fails to occur, but an external mechanism can provide an explanation for why the Trial failed, the Trial is declared *explained*. An example of an explained Trial is one in which the event fails to occur during the predicted time window, but instead appears shortly before or after that window.

The Predictor keeps track of its short- and long-term reliability by recording the number of successful, explained and failed Trials it has generated. We'll next see how this allows Predictors, through a process of reinforcement, to learn about causality on the job.

3.3.3 Learning Causality on the Job

Although some Predictors might be built offline, thus representing apparent temporal causality relationships the creature knows a priori, much of this knowledge must be learned during the creature's lifetime.

1. Concern yourself with interesting things.

An immediate challenge for anything but the most trivial of systems is the tremendous size of the perceptual state-space. Each stimulus might be considered another dimension of a massively multidimensional space that is probably only sparsely populated with areas of perceptual interest. Thus the first thing we must do is concern the learning mechanism with only the most interesting things. To do this, we add a salience filter between perception and action selection. Two heuristics determine whether a stimulus passes through the barrier: it can be interesting on the basis of its *novelty* (rarely perceived), or *inherent salience* (e.g. a loud bang).

2. Explain the Unexpected (with new Predictors)

Learning is prompted by the creature's inability to predict changes to the stimuli it considers interesting. In terms of the action selection mechanism described in 3.2, if the *React* operation is unable to find a Predictor that explains a salient stimulus onset, it is provided with an opportunity to generate a new explanation.

Explanation generation is guided by salient events that are temporally proximate to the unexpected stimulus. Recent Perception events on the TimeLine provide a convenient collection of all such candidates. To generate the appropriate Predictor, we need simply identify the stimulus (or group of stimuli) that seems the most likely explanation for the appearance of the unexplained stimulus.

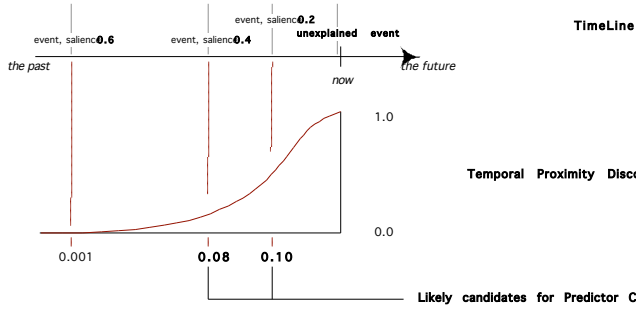

Figure 7: Selecting a likely Predictor context.

As illustrated in Figure 7, the explanation generator chooses a likely explanation that is both salient and temporally proximate to the unexplained stimulus. (Recall this may be some component of self-action, or an external perception.) It then builds a Predictor of the unexplained stimulus with this explanation as its Predictor Context. Although the particulars of the function used to select an explanation are unimportant, its probabilistic nature is crucial. The length of the Predicted Interval recorded in a new Predictor is equal to the perceived length of the time between the selected explanation and the unexplained event. When we have observed additional trials, the recorded interval is allowed to drift toward the average interval length.

3. Refine Predictors by tracking their reliability

If we find that a Predictor is reliable, our confidence in its predictive power will increase. On the other hand, if the Predictor is unreliable, we may either declare it invalid, or choose to refine it. One way to note changes in a Predictor's reliability is to detect differences between its recent and long term reliability.

The distinction between periodicity and probability is also important. A Predictor able to ideally predict a periodic reliability schedule (e.g. "an event will occur on every third trial") also requires a periodic function detector (like [17]). Note that a Predictor that expects an event to occur once every four times its context is observed causes quite different expectations than does a Predictor believed to be valid twenty-five percent of the time on a fixed ratio schedule. The former will generate a *high-confidence expectation every fourth time* the Predictor Context is observed; the latter will generate a *low-confidence expectation every time* the Context is observed.

A very simple metric for the long-term reliability of a Predictor is $R_{predictor}=(s_T+e_T)/(s_T+e_T+f_T)$, where s_T is the successful trial count, e_T the explained Trial count, f_T the failed Trial count. The Short-term reliability can be computed similarly by taking into account only several of the most recent Trials.

We guide the refining of a Predictor by determining the reliability and frequency of salient stimuli that are observed at the onset of its Trials. If a particular stimulus is both *salient* and *reliable* (often present at the start of successful trials, and often *not* present during failures), then that stimulus is a candidate.

If $s_\_$ denotes the number of times a stimulus _ was present during a successful trial, and $f_\_$ denotes the number of times _ was present during a failure, then the equation $R_\_=s_\_(f_T-f_\_)/(s_\_f_T+1)$ satisfies these features. The first factor $(s_\_/s_T)$ provides the ratio of successful trials in which the stimulus was present; the second factor $(f_T-f_\_/f_T)$ provides the ratio of unsuccessful trials in which the stimulus was not present. The additive term in the denominator prevents division by zero before we have at least one successful trial and one failed trial. Thus $R_\_$ increases as the stimulus is present in successful trials, and decreases as the stimulus is present in unsuccessful trials.

3.3.4 The representation of action reflects causality

We now have a representation for Prediction, but it will only be useful to the creature if the action selection mechanism can take advantage of apparent temporal causality knowledge. Until now, our discussion of action selection has been rather general. We have implemented and tested the Predictor mechanism and concepts presented here by augmenting the *ActionTuple,* the fundamental representation of action originally proposed by Blumberg (see [4]).

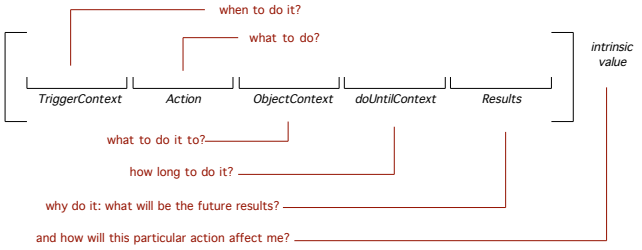

Figure 8: Anatomy of an (augmented) ActionTuple.

As seen in Figure 8, the ActionTuple encapsulates the concepts of trigger, action, object, doUntil, and in this new formulation, results. The TriggerContext indicates external conditions that must be met in order for the ActionTuple to be activated.

("When should I do it?") The Action represents what the creature should do if the ActionTuple is active. *("What should I do?")* The ObjectContext describes necessary conditions on the things to which that Action can be applied. *("What should I do it to?")* The doUntilContext describes the conditions that cause the ActionTuple to deactivate. *("How long should I do it?")* The Results slot contains Predictors, as described in the previous Section, each of which predicts that when the Tuple is activated, an event will occur after an interval with a certain probability. *("What will this cause?")* The Intrinsic Value is a multi-dimensional value (with the same dimensions as the DriveVector – see S3.1) that describes the Tuple's perceived effect on the creature's Drives. *("How will this help?")*

In Section 3.3, we indicated that each Predictor has a corresponding Predictor Context that determines when it generates expectations. We now see that the Predictors found in an ActionTuple's Results slot inherit their Predictor Context from the ActionTuple in which they are found. The TriggerContext, Action and ObjectContext slots conveniently denote external context, self-action, and the target of that action.

Intrinsic value is provided as a fixed value for some ActionTuples, which we refer to as consummatory ActionTuples. These are *reinforcers* (when they have high-magnitide, negative intrinsic values which suggest the satisfying of drives) and also *punishment* (large, positive intrinsic values which suggest an increase in drives.)

Although performing a particular action may not affect the creature's drives, an ActionTuple's predicted Results may change the world in a way that would facilitate satisfying drives in the future. Thus the utility of an ActionTuple is defined by more than just its intrinsic value. The perceived value of an ActionTuple is a function of its intrinsic value, and the perceived value of ActionTuples that could be potentially activated due to causality information contained in the Results slot. Importantly, a predicted Result is valuable if and only if, through some causal chain, it will help satisfy a currently unsatisfied prerequisite of a consummatory Tuple. Thus the perceived value of an ActionTuple changes as our needs change, *and* as the perceived external conditions in the world change.The perceived value of an ActionTuple is calculated

where $v_i(t)$ is

the intrinsic value of Tuple i, R_m the reliability of each associated Predictor, $pv_n(t)$ the perceived value of each Tuple facilitated by that Predictor, and *k* a discount factor. In this implementation, this equation uses a max recursive depth of 4.

3.3.5 Changing ideas about causality

We'll use an example to show how ActionTuples and the mechanism described above can represent and learn an apparent temporal causality relationship. Consider an experiment wherein a dog is conditioned to salivate upon hearing a bell ring, because the bell provides a reliable predictor of the appearance of steak.

We begin with the assumption that the dog has the inherent idea that consuming steak will reduce his hunger drive. We construct the consummatory ActionTuple that represents this relationship (assuming the animal has only two drives, hunger and sex).

The consummatory act of eating the food is represented by the ActionTuple depicted on the bottom of Figure 9: with a null TriggerContext (meaning no external conditions need to be met), the eat Action in the Action slot, the foodShape stimulus as an ObjectContext (meaning the action must be performed on food, and thus can't be performed unless food is present), and the notion "until consumed" in the doUntilContext. The intrinsic value [-10,0] indicates that the creature's hunger drive will be reduced if the Tuple is activated. If the creature has a sufficiently high hunger drive, an action selection mechanism like the one described in S3.2 would be inclined to activate this Tuple when the creature perceives food.

During this experimental procedure, the dog will be presented with two salient stimuli: the sound of the bell, and the appearance of a steak. In her attempts to explain these unexplained stimuli, the dog will, after a time, come to the idea that the sound of the bell is reliably followed by the appearance of food. This bell-predicts-steak notion is represented by the Tuple shown at the top of the figure. The TriggerContext for this Tuple is the bellSound, the ObjectContext null, the Action null, and the doUntilContext null. The Results slot contains the Predictor indicating that something with the foodShape property will appear in a few seconds with 33% reliability. Although the intrinsic value of the "hearing a bell" Tuple is null (zero), the concept of perceived value makes its activation seem like a good thing to the dog. It indicates that the activation of this Tuple reliably leads to the activation of another Tuple that will satisfy the hunger drive.

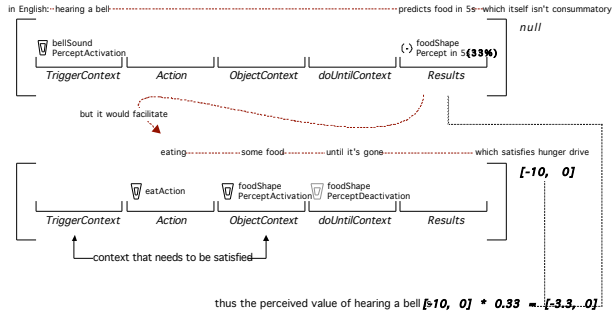

Figure 9: Perceived Value makes hearing the bell good.

The perceived value of the "hearing the bell" Tuple is calculated by the sum of its intrinsic value, and the intrinsic values of the Tuples it facilitates multiplied by a discount factor. The discount factor for each term is a function of the probability that the required stimulus will appear. In this example, the dog is conditioned that food will appear on average every one in three trials. Thus the discount factor is 1/3, and since the perceived effect on the hunger drive of eating the predicted food is -10, the perceived value of the "hearing the bell" ActionTuple is -10/3.

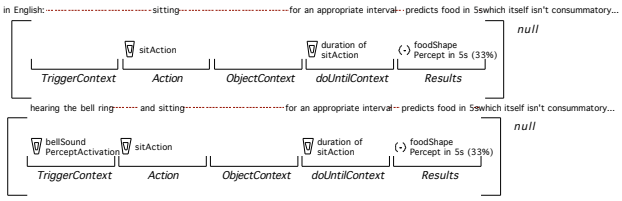

Figure 10: Self-Action Variation of the experiment.

A variation of this experiment (Figure 10) involves only providing the dog with a reinforcer when it sits down after the bell rings. In this case, the dog may begin with the hypothesis (top of figure) that simply sitting down predicts the treat. Because some of the trials will be reinforced and others will not, the Predictor will eventually realize that the bell sound reliably predicts the trial's outcome. Thus, a new Tuple (bottom of figure) will be created. The Action is still "sit", but now the bell sound has been added to the TriggerContext. The Results contain a Predictor predicting the foodShape's onset in a few seconds. Thus, predictions can, but do not have to, involve self-action.

3.4 Causality, Affect and Reward Markers

In addition to allowing an agent to select actions, apparent temporal causality also provides feedback that can inform the creature's motivational and affective state.The utility of something in the world can also be interpreted as a creature's affective stance towards something. The creature can use its affective stance toward a stimulus to generate appropriate reactions to its onset – and predictions of its impending onset. A creature can also use the affective stance to determine whether or not it wishes to encourage the onset of a stimulus. Thus an interesting effect of the DriveVector approach is that a creature's emotional memories of some context are affected by its current needs.

A creature's motivational state also affects the action selection mechanism's propensity to explore rather than exploit. It thus has an indirect but important effect on learning by altering the rate at which the action selection mechanism generates and refines Predictors.

Our ability to compute the affective value of a stimulus offers us flexibility in the way we produce reward markers for machine learning algorithms elsewhere in the system. Many such algorithms (e.g. the one that drives acoustic category formation in our acoustic pattern matcher) employ a reward marker (and sometimes a punishment marker) to inform the classifier of the results of a recent classification. The fundamental question is: which stimuli constitute reward markers? An obvious answer is a stimulus that indicates the appearance of a reinforcer like food. But there also may also be times when we can predict the *impending* onset of a reinforcer with sufficiently high confidence that we can post the reward marker before the reinforcer actually appears. Here this occurs at the moment when we can first predict, with confidence above a threshold, the future appearance of all the stimuli necessary to activate a consummatory ActionTuple.

4. Results

The architecture has been demonstrated capable of reproducing a wide variety of conditioning phenomena, such as those described in Section 3, as well as providing a robust basis for an implementation of the clicker training paradigm [25]. From an agents perspective, this work provides a model for action selection and learning that integrates apparent temporal causality into the action selection mechanism of a complete virtual creature.

The time scale invariance suggested by SET provided an elegant representation of internal timing in the Predictor representation. Because they record the perceived interval between two stimuli and use _-thresholds to generate windows in which stimuli are predicted, Predictors are able to represent short intervals with high precision, and exhibit plausible uncertainty for longer intervals. The implementation approximates RET by employing the principle of parsimony when a creature builds reliable predictors to explain its world, learning only the simplest explanation for how its world works.

5. Related Work

Tu and Terzopoulos's physically-based artificial fish model incorporates perception and action selection [19]. Perlin's Improv system is designed to create interactive actors [20]. Damasio's somatic markers are the precursor to our drive-based value attribution [21]. The importance of considering perception and learning together was emphasized by Barlow [24]. Maes and Drescher discuss reliability [26], [27]. Allen's work integrates temporal reasoning into a planning system [15]. deKleer introduces and describes causal theories [14]. Further discussion of the application of causality is found in [22]. Pearl's discussion of "explaining away" events [13] influenced the Explain operation. See Moray for more on the structure of mental models [16]. The structure of this work is largely inspired by Blumberg's Ph.D. thesis [3]. Gallistel and Gibbon's timing model [6] contrasts sharply with the standard model of conditioning formalized by Rescorla & Wagner [23].

6. ACKNOWLEDGMENTS

Our thanks to the members of the Synthetic Characters Group, Whitman Richards and Randy Gallistel for invaluable insight.

7. REFERENCES

[1] Gallistel, C. R. (1990). The Organization of Learning. Cambridge, MA, Bradford Books / MIT Press.

[2] Brooks, R. A. (1991b). "Intelligence Without Representation." Artificial Intelligence Journal **47**: 139-159.

[3] Blumberg, B. M. (1996). Old Tricks, New Dogs: Ethology and Interactive Creatures. Media Lab. Cambridge, MIT.

[4] Isla, D. A., R. C. Burke, et al. (2001). A Layered Brain Architecture for Synthetic Characters. IJCAI, Seattle.

[5] Wilkes, G. (1994). Behavior Sampler, C&T Publishing.

[6] Gallistel, C. R. and J. Gibbon (2000). "Time, Rate and Conditioning." Psychological Review **107**: 289-344.

[7] Thorndike, E. (1911). Animal Intelligence. Darien, Hafne.

[8] Yoon, S.-Y., B. M. Blumberg, et al. (2000). Motivation Driven Learning for Interactive Synthetic Characters. AA 2000.

[9] Russell, J. (1980). "A circumplex model of affect." Journal of Personality and Social Psychology **39**: 1161-1178.

[10] Ekman, P. (1982). Emotion in the Human Face. Cambridge, UK, Cambridge University Press.

[11] Brooks, R. A. (1991a). Intelligence without Reason, Computers and Thought lecture. IJCAI-91, Sidney, Australia.

[12] Burke, R. C., D. A. Isla, et al. (2001). Creature Smarts: The Art and Architecture of a Virtual Brain. Game Developers Conference 2001.

[13] Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. San Mateo, California, Morgan Kaufmann Publishers.

[14] deKleer, J. (1986). "An Assumption-based TMS." Artificial Intelligence Journal **28**(2): 127-162.

[15] Allen, J. F. (1991). Planning as Temporal Reasoning. The Second International Conference on Principles of Knowledge Representation and Reasoning, Morgan Kaufmann.

[16] Moray, N. (1990). "A lattice theory approach to the structure of mental models." Phil. Trans. R. Soc. Lond. **B**(327): 577-583.

[17] Aittokallio, T., M. Gyllenberg, et al. (2000). Testing for Periodicity of Signals: An Application to Detect Partial Upper Airway Obstruction during Sleep, Turku Centre for Computer Science.

[18] Spier, E. (1997). From Reactive Behaviour to Adaptive Behaviour: Motivational Models for Behavior in Animals and Robots. Oxford, Oxford University: 99.

[19] Tu, X. and D. Terzopoulos (1994). Artificial Fishes: Physics, Location, Perception, Behavior. Siggraph.

[20] Perlin, K. and A. Goldberg (1996). "Improv: A System for Scripting Interactive Actors in Virtual Worlds." Computer Graphics 29(3).

[21] Damasio, A. (1995). Descarte's Error, Harvard University Press.

[22] Iwasaki, Y. and H. Simon (1986). "Causality in Device Behavior." Artificial Intelligence Journal 29(1): 3-32.

[23] Wagner, A. R. and R. A. Rescorla (1972). Inhibition in Pavlovian conditioning: Application of a theory. Inhibition and Learning. R. A. Boakes and M. S. Halliday. London, Academic Press.

[24] Barlow, H. (1990). "Conditions for Versatile Learning, Helmholtz's Unconscious Inference, and the Task of Perception." Vision Research 30(11): 1561-71.

[25] Burke, R.C. (2001). It's About Time: Temporal Representations for Synthetic Characters. MS. Thesis, The Media Lab, MIT.

[26] Maes, P. (1989). The Dynamics of Action Selection. IJCAI, Detroit, Morgan Kaufmann.

[27] Drescher, G. L. (1991). Made-up minds: a constructivist approach to artificial intelligence. Cambridge, Mass., MIT Press.