Domains on Classical States
Final Project for COMP-627

Jordan Frank

April 23, 2007

Abstract

In this paper, we will investigate two partial orders on classical states Δ^n that induce a domain theoretic structure. In particular, we will look at the *Bayesian order* presented by Coecke and Martin (2002) and the *implicative order* presented by Martin (2004). For both of the domains that arise from these orderings on the classical states, Shannon entropy is a measurement in the domain theoretic sense. We will then investigate the behaviour of a simple learning algorithm for stochastic learning automata, which has previously been studied using domain theory (Edalat, 1995), in terms of how the states of the system change with respect to the domains induced by these partial orders.

1 Introduction

The classical states are a natural way of representing the state of many systems in many domains, from learning systems to quantum computing systems, but not a lot of work has been done to study the domain theoretic properties of classical states. In this paper, we will give an overview of some of the work that has been done by a few authors to assign order to the states. This is an interesting idea, assigning order to states that are commonly used to represent disorder, and we will look at the structures that arise from these orders. What is interesting is that in both of the orders that we look at, Shannon entropy, a standard measure for the uncertainty in a state, provides a quantitative expression of the qualitative notion of a state being more informative than another state (ie. a state being greater than another state in the given order). We will assume that the reader is familiar with the basic definitions from Domain theory that would be covered in a basic introductory course on the subject, and for the sake of brevity we will not reproduce them here.

We will begin by introducing the classical states, and then discuss the notion of measurement in terms of domain theory. Then we will look at two partial orders on classical states, the Bayesian order and the implicative order, and we will discuss the motivation behind the two, and the domain theoretic structure that arises when we apply the order to the classical states. We then analyze a simple learning algorithm for stochastic learning automata, the linear reward-penalty scheme, in terms of how the states progress with respect to the two orders, as the system learns. We end with a brief conclusion, and a discussion of future directions for study.

2 Domains of classical states

In this section, we will discuss two partial orders on the set of classical states, both of which induce a domain theoretic structure. First we will define what we mean by the classical states and what we mean by a measurement on a domain, thus setting up the following sections which will define first the Bayesian order from Coecke and Martin (2002), followed by the implicative order from Martin (2004).

2.1 Classical states

If we have an event in which one of n different outcomes is possible, then we can represent the information an observer would have about the result of the event by an n-tuple $(x_1, \dots, x_n)$, where x_i represents the belief, or probability, that outcome i will occur. We call these the classical states, which we define formally as

$$\Delta^n := \{(x_1, \dots, x_n) \in [0,1]^n : \sum_{i=1}^{n} x_i = 1\},$$

for $n \geq 2$. The set of classical states is equivalent to the n-simplex.

For notational convenience, we define

$$x^+ = max_{1 \leq i \leq n} x_i$$

and

$$x^- = max_{1 \leq i \leq n} x_i.$$

Supposing that we have no knowledge at all about the system, and no reason to believe a priori that any outcome is more likely than any other. This would be represented as the unique state $x \in \Delta^n$ such that $x^+ = x^- = 1/n$, which is referred to as the *completely mixed state*, or $\perp$. We call a state $x \in \Delta^n$ *pure* if $x^+ = 1$, and Δ^n has exactly n pure states which we denote $\{e_i : 1 \leq i \leq n\}$, and these correspond to the vertices of the n-simplex.

We also consider a special subset of the classical states, $\Lambda^n \subseteq \Delta^n$, the *monotone* decreasing states, also referred to as just the monotone states, defined as

$$\Lambda^n := \{x \in \Delta^n : (\forall i < n) x_i \geq x_{i+1}\}.$$

We will be looking at finite states, that is states with a finite number of elements, though the case of infinite states is presented by Mashburn (2006). Mashburn defines an order relation on infinite states, but the structure that is induced does not have a bottom element nor does it have Shannon entropy as a measurement, and so we will not discuss it any further. In Martin (2006), an order on the monotone states is given which is equivalent to majorization, but as there is no clear way how to extend this ordering from Λ^n to Δ^n while retaining some of the important domain theoretic properties such as continuity, we simply remark that Λ^n with majorization as the order relation is a continuous dcpo with a least element, and so does have some remarkable properties, but since only want to focus on orders on Δ^n, it lies outside of the scope of this paper.

An ideal order on the classical states would have the completely mixed state as the bottom element, and the pure states as the maximal elements. *Shannon entropy*

$$\mu x = -\sum_{i=1}^{n} x_i \log x_i$$

gives us a quantitative measure of the amount of uncertainty that a state possesses. Similarly, if we have two states x, y in a domain with an order relation $\sqsubseteq$, then asserting that $x \sqsubseteq y$ translates to the statement that y is more informative than x. Therefore it would also be ideal if the order relation that we define on the classical states would submit to Shannon entropy as a measurement in the domain theoretic sense. In particular, as states become more informative, uncertainty should decrease:

$$x \sqsubseteq y \implies \mu x \geq \mu y.$$

It is clear from the equation for uncertainty that the pure states, or the maximal elements, will have Shannon entropy 0, and the completely mixed state, or bottom element, will have the highest possible Shannon entropy, $\log n$.

Finally, we would like our order relation to respect *the mixing law*:

$$x \sqsubseteq y \;\&\; p \in [0,1] \implies x \sqsubseteq (1-p)x + py \sqsubseteq y.$$

The mixing law says that if y is more informative than x, then any mixture of the two states should be more informative than x, but less informative than y.

2.2 Measurement in Domain Theory

The notion of a *measurement* in the context of domain theory is introduced by Martin (2000). Since one of the desired properties of our domain on classical states is to have Shannon entropy as a measurement, we will need to cover what is meant by a measurement in domain theory.

We use the set $[0,\infty)^*$ to denote the nonnegative reals in their opposite order, and we make use of the following two definitions from Martin (2000).

Definition 2.1. A Scott continuous map $\mu : D \to [0,\infty)^*$ on a continuous dcpo D *induces the Scott topology near* $X \subseteq D$ if for all Scott open sets $U \subseteq D$ and for any $x \in X$,

$$x \in U \implies (\exists \varepsilon > 0)\, x \in \mu_\varepsilon(x) \subseteq U,$$

where $\mu_\varepsilon(x) = \{y \in D : y \sqsubseteq x \;\&\; |\mu x - \mu y| < \varepsilon\}$. This is written $\mu \to \sigma_X$.

Definition 2.2. A *measurement* on a domain D is a Scott continuous mapping $\mu : D \to [0,\infty)^*$ with $\mu \to \sigma_{\ker \mu}$ where $\ker \mu = \{x \in D : \mu x = 0\}$.

Measurement, then, is a process by which we can express the qualitative notion captured by a domain $(D, \sqsubseteq)$ in a quantitative expression. Martin gives a number of extra properties of measurements, and uses the idea of measurement to prove two very interesting fixed point theorems, but for the purpose of this paper we will only be using these two definitions.

2.3 Bayesian order

The first order on classical states that we will look at is the *Bayesian order* presented by Coecke and Martin (2002). The Bayesian order is developed for use with both classical and quantum states, but we will be focusing on just the classical states. The important definitions, theorems, and lemmas from Coecke and Martin (2002) have been transcribed in Appendix A in order to provide a more concise overview of the contributions from this rather lengthy, but extremely well written, paper.

The authors introduce the order using an amusing example to explain the rationale behind the Bayesian order, which we will modify slightly for our purposes. A grad student is trying to find an interesting topic to write a term paper about, and his supervisor hands him a stack of n papers to read. In fact, only one of the papers actually contains a topic that would be interesting enough to write a paper on. Based on the titles, the grad student can represent his belief in the probability that each paper will be interesting as a classical state $x \in \Delta^n$. Now, after the student reads paper i and discovers that it contains nothing of interest, he can update his belief in two ways: he can set the value $x_i = 0$ and add a small value to each of the other x_j's to keep the sum equal to 1, or he can form a new classical state $x' \in \Delta^{n-1}$, completely removing the entry for paper i. Once the student discovers that a paper is not interesting, there is no chance that he will later come back to that paper, and so we will choose the latter method because it removes the entry that we will no longer need to consider, forming a simpler representation. We can define a projection that incorporates this idea of collapsing the i^{th} outcome by the partial map $p_i : \Delta^n \rightharpoonup \Delta^{n-1}$ given by

$$p_i(x) = \frac{1}{1 - x_i}(x_1, \ldots, \widehat{x_i}, \ldots, x_n)$$

for $1 \leq i \leq n$ and $0 \leq x_i < 1$, where the $\widehat{\cdot}$ operator denotes the removal of an element from the state. This projection can only represent the notion of finding out that a certain event is *not* the correct one, and so it is only defined on $\mathrm{dom}(p_i) = \Delta^n \setminus \{e_i\}$. The case where the student reads the interesting paper is not interesting to us, since it essentially ends the whole process.

The authors begin by defining an order relation $\sqsubseteq$ on the classical two states, Δ^2 such that for $x, y \in \Delta^2$,

$$(x_1, x_2) \sqsubseteq (y_1, y_2) \iff (y_1 \leq x_1 \leq 1/2) \text{ or } (1/2 \leq x_1 \leq y_1).$$

The authors then go onto prove that the given order is the unique partial order on Δ^2 that both satisfies the mixing law and has a bottom element $\perp = (1/2, 1/2)$. The authors then go on to define the order on Δ^n recursively, such that for $n \geq 2$, $x, y \in \Delta^n$

$$x \sqsubseteq y \iff (\forall i)(x, y \in \mathrm{dom}(p_i) \implies p_i(x) \sqsubseteq p_i(y)),$$

where i ranges over the set $\{1, \ldots, n\}$. This essentially says that if observer A has more information than observer B about the true outcome of an event, then after learning that any single outcome is not the correct one, A should still have more information than B. Though it is not explicitly stated in the paper, we believe that this order is called the Bayesian order because it is loosely based on the idea of conditional probability in that an state x is more informative than a state y if and only x is still more informative than y given any possible single observation.

Next, the authors prove that with the Bayesian order, the classical states Δ^n are a dcpo, and moreover they prove an interesting property of the classical states, that every directed subset contains an increasing sequence with the same supremum, and so we can always replace directed sets with increasing sequences, which greatly simplifies any argument that involves directed sets.

The problem with how we've defined the Bayesian order is that it involves both a recursive quantifier as well as recursion, and so it makes the process of determining whether two states are related a fairly involved process. In order to simplify the definition of the ordering relation, the authors introduce the symmetric group

$$S(n) = \{\sigma | \sigma : \{1, \ldots, n\} \simeq \{1, \ldots, n\}\}$$

of bijections on the set $\{1, \ldots, n\}$, referred to as *permutations* or *symmetries*. For any state $x \in \Delta^n$, there exists a permutation σ such that $x \cdot \sigma \in \Lambda^n$, that is to say that for any state, there is a permutation that makes the state monotone. With this in mind, we can then we can redefine our ordering relation $\sqsubseteq$ such that for $x, y \in \Delta^n$, we have $x \sqsubseteq y$ iff there is a permutation σ of $\{1, \ldots, n\}$ such that $x \cdot \sigma$ and $y \cdot \sigma$ are monotone and

$$(x \cdot \sigma)_i (y \cdot \sigma)_{i+1} \leq (x \cdot \sigma)_{i+1} (y \cdot \sigma)_i$$

for all i with $1 \leq i < n$.

This new definition gives us an easier way to determine the relation between two states. First we need to determine if there exists a permutation σ that we can apply to both x and y to make them monotone, and then if there is, then we simply compare each pair of subsequent elements as described above. This highlights the fact that for two states $x, y \in \Delta^n$ to be comparable, the indices i such that $x_i = x^+$ must be the same as the indices j such that $y_j = y^+$, and similarly the indices i such that $x_i = x^-$ must be the same as the indices j such that $y_j = y^-$, otherwise we would not be able to find a permutation σ such that $x \cdot \sigma$ and $y \cdot \sigma$ are both monotone.

The authors proceed to discuss what they call *symmetric* functions, which are functions $f : \Delta^n \to E$ where E is a dcpo and for all $\sigma \in S(n)$, $f(x \cdot \sigma) = f(x)$. It just so happens that Shannon entropy is a symmetric function. More importantly, the natural retraction $r : \Delta^n \to \Lambda^n$ is symmetric, and so for problems that involve classical states, we can first focus on solving the problem for Λ^n with the knowledge that the solution will naturally extend to Δ^n in general.

Now that we have established that the classical states with the Bayesian order is a dcpo, we look at the notion of approximation. It is at this point that the authors deviate from the standard definitions used in domain theory. The definition for the "way below" relation is slightly different, and for the purpose of this paper we will call this relation "weakly way below", which we denote as $\ll_w$ and define as follows: Let D be a dcpo. For $x, y \in D$, we write $x \ll_w y$ iff for all directed sets $S \subseteq D$,

$$y = \bigsqcup S \implies (\exists s \in S)\, x \sqsubseteq s.$$

Note that this differs from the usual definition of way below in that it has an equality rather than an inequality. The standard definition is that for a dcpo D, for $x, y \in D$

we write $x \ll y$ iff for all directed sets $S \subseteq D$,

$$y \leq \bigsqcup S \implies (\exists s \in S)\, x \sqsubseteq s.$$

While the difference between the two definitions seems minute, it actually has quite a drastic effect on the *approximations* of an element x of a dcpo D, which we define as

$$\twoheaddownarrow_w x := \{y \in D : y \ll_w x\},$$

which is again similar to the standard definition of $\twoheaddownarrow x$ which uses the standard way below relation instead of the weak version. We will delve further into the difference in a moment, but we need a few more definitions in our arsenal.

Similar to the definition of a *continuous* dcpo, where $\twoheaddownarrow x$ is directed with supremum x for all $x \in D$, we call a dcpo *exact* if $\twoheaddownarrow_w x$ is directed with supremum x for all $x \in D$. Clearly a continuous dcpo is also an exact dcpo, since $x \ll y \implies x \ll_w y$, and the two notions also coincide on maximal elements, since if $e \in D$ is a maximal element, and S is a directed set S with supremum $\bigsqcup S \geq e$, then $\bigsqcup S = e$, and so $\twoheaddownarrow_w e = \twoheaddownarrow e$.

The straight line path from x to y is

$$\pi_{xy}(t) = (1-t)x + ty,$$

where t is in the range $0 \leq t \leq 1$. It is clear that $\pi_{xy}(0) = x$ and $\pi_{xy}(1) = y$, and the authors show that π_{xy} is Scott continuous iff $x \sqsubseteq y$. The authors then show that for any $x \in \Delta^n$, $\pi_{\perp x}(t) \ll_w x$ for $t < 1$. Therefore, for every $x \in \Delta^n$, the set $\twoheaddownarrow_w x$ contains a directed set with supremum x. Now, for any dcpo D, for each $x \in D$, the set $\twoheaddownarrow x$ is directed with supremum x iff it contains a directed set with supremum x, and so it follows that Δ^n is exact.

Next, the authors prove a proposition that brings to light the reason why they use the modified version of the way below relation. The proposition is that for $n \geq 2$, for all $x \in \Delta^n$, $x \ll_w e_i$ iff $x = \pi_{\perp e_i}(t)$ for some $t < 1$. This means that any element that is weakly way below a maximal element e_i must lie on the straight line path between $\perp$ and e_i. The authors do not go into much more detail as to how this restricts the domain, but for the purpose of this paper we will investigate a bit further.

The main difference between an exact dcpo and a continuous dcpo is embodied in the idea of *context*. In a continuous domain D with order relation $\leq$ and elements $x, y, z \in D$, we have that $x \ll y \leq z \implies x \ll z$. In an exact domain, this relationship only holds if the statements $y \leq z$ and $x \ll y$ are being made in the same context. The authors show that if x, y, and z are all regarded as necessary for a single state, that is if $\uparrow z \neq \emptyset$, then we can say that $x \ll z$, but otherwise it may not necessarily be true.

Considering this, it becomes apparent why the modified definition of the way below relation is used. If the authors had used the standard way below definition, then most states would not have approximations. For example, suppose we have the state $y = (1/2, 1/2, 0) \in \Delta^3$, and suppose that $\perp \neq x \ll y$. We have that $y \sqsubseteq e_i$, and so $x \ll e_i$ which means that $x \ll_w e_i$. Therefore, from the proposition above, we have that x must lie on the straight line path between $\perp$ and e_i. But the Bayesian order has a degenerative property, that is that if $x \sqsubseteq y$, then $y_i = y_j > 0 \implies x_i = x_j$, and so since $x \sqsubseteq y$ and $y_1 = y_2 = 1/2 > 0$, we must have that $x_1 = x_2$. The only state with $x_1 = x_2$ that is on the straight line path between $\perp$ and e_1 is $\perp$, which contradicts our assumption that $x \neq \perp$. Therefore the only element that is way below y by the standard definition is $\perp$.

Simply having one example is not enough to claim that most states do not have approximations, and towards making this paper more than just a summary of other peoples' work, we will now prove that the classical states with the Bayesian order is not continuous. First we need a trivial proposition. We cannot recall whether it was covered in the course, but since it is so simple, we will include it here.

Proposition 2.3. *Let D be a dcpo. For $x, y, z \in D$, if $x \ll y$ and $y \leq z$, then $x \ll z$.*

Proof. Since $x \ll y$, every directed set $S \subseteq D$ with $\bigsqcup S \geq y$ has an element $s \in S$ such that $x \leq s$. Then any directed set $S \subseteq D$ with $\bigsqcup S \geq z \geq y$ must have an element $s \in S$ such that $x \leq s$, and so $x \ll z$.[1] □

We are now ready to prove our claim, noting that since we included the pertinent results from Coecke and Martin (2002) in Appendix A, we will be making reference to the results as they are indexed in the Appendix, and not in the original paper.

Proposition 2.4. *Let Δ^n denote the classical states with the Bayesian order. Δ^n is not a continuous dcpo.*

Proof. We will prove that for any element $x \in \Delta^n$ such that x does not lie on a straight line path between $\perp$ and some maximal element e_i, $\bigsqcup \twoheaddownarrow x \neq x$. Since most elements of Δ^n do not lie on one of these paths, it follows that Δ^n is not continuous.

Let $x \in \Delta^n$ be an element that does not lie on a straight line path between $\perp$ and any maximal element. Let $e_i \in \Delta^n$ be a maximal element such that $x \leq e_i$, and let $a \in \Delta^n$ be an element such that that $a \ll x$. By Proposition 2.3, we have that $a \ll e_i$. Now, since e_i is a maximal element, $a \ll e_i \iff a \ll_w e_i$, and so Proposition A.34(i) tells us that a must lie on the straight line path between $\perp$ and e_i. This straight line path is closed under directed suprema, and so the sup of any elements that are on the line must also be on the line. Therefore, since x is not on the line, $x \neq \bigsqcup\{a \in \Delta^n : a \ll x\} = \bigsqcup \twoheaddownarrow x$. □

In a private correspondence, Keye Martin informed us that it was in fact possible to prove that if we remove the boundary states $\mathcal{B} \subseteq \Delta^n$, that is those states that have elements equal to 0, such as the maximal elements or y in the previous example, then the result is in fact a continuous dcpo, but time constraints prevented us from investigating this further, other than noting that a dcpo D is continuous if it is exact and $x \ll y \sqsubseteq z \implies x \ll z$ for all $x, y, z \in D$. Since this implication will hold provided $\twoheaduparrow z \neq \emptyset$, we could probably show that Δ^n is still an exact dcpo if we remove the boundary states, and any element $z \in \Delta^n \setminus \mathcal{B}$ will approximate some element $z' \in \Delta^n \setminus \mathcal{B}$ (ie. $z \ll z'$, and so $\twoheaduparrow z \neq \emptyset$).

Finally, the authors show that Shannon entropy μ on classical states is Scott continuous on Δ^n, and is a measurement in the domain theoretic sense. Since it is monotone, it must be the case that for $x, y \in \Delta^n$, $x \sqsubseteq y \implies \mu x \geq \mu y$, and so Shannon entropy gives us a quantitative expression of what we mean by the qualitative statement $x \sqsubseteq y$.

[1] This probably should have just been described as being obvious.

2.4 Implicative order

Given the shortcomings of the Bayesian order, namely that the induced dcpo is not continuous, and so not a domain in the technical sense, it was only a matter of time that an order on the classical states was found that would give a continuous domain. This order is called the implicative order, and is presented by Martin (2004). In this section, the relation $\sqsubseteq$ will represent the implicative order, which we define as follows: For $x, y \in \Delta^n$, $x \sqsubseteq y$ iff for all $i \in \{1, \ldots, n\}$, $x_i \geq y_i$ or $x_i = x^+$. The authors then show that Δ^n with the implicative order is a continuous dcpo.

According to the authors, it is important to understand why this order is called the implicative order, and so we will now discuss this matter. We say that an element x of a dcpo D is *irreducible* when

$$\bigwedge(\uparrow x \cap \max(D)) = x$$

and we denote the set of all irreducible elements in D as $\mathrm{Ir}(D)$. The authors then show that for any $n \geq 2$, the set of irreducible elements of Δ^n is isomorphic to $\mathcal{P}\{1, \ldots, n\} \setminus \{\emptyset\}$ ordered by reverse inclusion. This means that the power set of $\{1, \ldots, n\}$ sits nicely in Δ^n, and each $A \in \mathcal{P}\{1, \ldots, n\} \setminus \{\emptyset\}$ corresponds to a classical state $x \in \mathrm{Ir}(\Delta^n)$ given by

$$x_i := \begin{cases} x^+ \text{ if } i \in A, \\ 0 \text{ otherwise.} \end{cases}$$

The authors then go on to explain how Δ^n is actually a continuous extension of $\mathcal{P}\{1, \ldots, n\}$. For a set $A \in \mathcal{P}\{1, \ldots, n\}$, we have a characteristic function $\chi_A : \{1, \ldots, n\} \to \{0, 1\}$ defined by

$$\chi_A(i) := \begin{cases} 1 \text{ if } i \in A, \\ 0 \text{ otherwise.} \end{cases}$$

Then if we take $\leq$ to be the pointwise order on function of the type $\{1, \ldots, n\} \to \{0, 1\}$, then we have

$$A \supseteq B \iff \chi_A \leq \chi_B.$$

Now we assign to a state $x \in \Delta^n$ a characteristic function $\chi_x : \{1, \ldots, n\} \to [0, 1]$ given by

$$\chi_x(i) := \begin{cases} 1 \text{ if } x_i = x^+, \\ x_i \text{ otherwise.} \end{cases}$$

These functions χ_x then extend our notion of the characteristic functions on elements of the power set to the elements of our continuous domain Δ^n, and they agree on the elements $\mathrm{Ir}(\Delta^n)$. The authors then prove that for $x, y \in \Delta^n$, $x \sqsubseteq y \iff \chi_x \geq \chi_y$, thus justifying the use of the term implication for this order. Finally, the authors prove that Δ^n is an ω-continuous dcpo with Shannon entropy as a measurement.

3 Stochastic Learning Automata

Stochastic learning automata, sometimes referred to as cellular automata, are learning systems that were originally studied in the field of mathematical psychology (Tsetlin, 1973), but later had a large impact in the field of control theory (Narendra and Thathachar, 1989). The standard framework is we have an automaton that exists in an unknown, random environment in which the automaton can issue one of a finite number of actions to the environment, which will respond with a signal indicating success or failure. The goal of the automaton is to choose actions that have the highest probability of success, and the automaton is said to learn when if it tends to increase its chance of success as it proceeds. Formally, we define the framework as the tuple $(\mathcal{A}, \mathcal{S}, \pi, \mathcal{B}, \mathcal{F})$ where

- $\mathcal{A}$ is the set of r actions
$$\mathcal{A} = \{1, \ldots, r\}.$$
- $\mathcal{S}$ is the set of internal states of the automaton where $\mathbf{s} \in \mathcal{S}$ is a probability vector
$$s = (s_1, \ldots, s_r), \quad \sum_{i=1}^{r} s_i = 1$$
where s_i denotes the estimate of the success probability of action i.
- π is the environment, and is also a probability vector
$$\pi = (\pi_1, \ldots, \pi_r), \quad \sum_{i=1}^{r} \pi_i = 1$$
where π_i represents the actual probability of action i succeeding. For the purpose of this paper, we will assume that the environment is stationary, and so the probabilities do not change over time.
- $\mathcal{B} = \{0, 1\}$ is the set of signals that the environment will issue in response to an action. Here 1 indicates success and 0 indicates failure.
- $\mathcal{F}$ is the set of continuous maps representing the learning functions
$$\mathcal{F} = \{f_e : S \to S | e \in E\}$$
where $E = \mathcal{A} \times \mathcal{B}$ is the set of pairs (i, b) representing an event in which the automaton issued action $i \in \mathcal{A}$ and received the response $b \in B$.

It can easily be seen then, that if the automaton is in the state $\mathbf{s}$, then the probability of an event $e = (i, b) \in E$ is given by

$$p_e(\mathbf{s}) = s_i(\pi_i b + (1 - \pi_i)(1 - b)).$$

If an automaton is in state $\mathbf{s}$ and the event $e = (i, b) \in E$ occurs, then the automaton updates its state to be $f_e(\mathbf{s})$. The dynamics of the system can then be described as follows: The automaton starts the iteration in state $\mathbf{s}$, observes the event $e = (i, b) \in E$ with probability $p_e(\mathbf{s})$, and updates its state

$$\mathbf{s} \leftarrow f_e(\mathbf{s}).$$

3.1 Linear reward-penalty scheme

A very basic learning algorithm for stochastic learning automata is the linear reward-penalty scheme. In this scheme, the automata observes an event $e = (i, b)$, and updates the state using the learning function

$$f_e(\mathbf{s}))_j = \begin{cases} s_j + \lambda(1 - s_j) \text{ if } j = i, \\ (1 - \lambda)s_j \text{ if } j \neq i \end{cases}$$

if $b = 1$, or the learning function

$$f_e(\mathbf{s}))_j = \begin{cases} (1 - \lambda)s_j \text{ if } j = i, \\ (1 - \lambda)s_j + \lambda/(r - 1) \text{ if } j \neq i \end{cases}$$

if $b = 0$, for $0 \leq j \leq r$ in both cases. $\lambda \in (0, 1)$ is the learning rate of the system.

The linear reward-penalty scheme for a binary-state automaton is modeled in terms of an iterated function system (IFS) on a probabilistic power domain by Edalat (1995). In the paper, the author shows that the limiting distribution of the states, that is the distribution that the states will eventually converge to, can be calculated as a generalized Riemann sum. This is an interesting result, and it is neat to see that the results of the domain theoretic analysis of the problem agree with the well-known results about the binary-state automaton, which employ classical measure theory.

Instead of focusing on the binary-state automaton, as so many others do, we will look at the behaviour of the linear reward-penalty scheme for an r-state automaton. It is clear that the set of states, $\mathcal{S}$, of the automaton are in fact our classical states, Δ^n, and so we will now look at the behaviour of the states of the automaton using the two orders on Δ^n that we studied in the previous section.

3.2 Analysis

When we set out to study the behaviour of the linear reward-penalty scheme with respect to the orders on classical states, we had hoped that the results would be far more interesting than those we actually ended up with. The problem is that if we consider the states as discrete probability distributions, then for two states to be comparable, they have to have their peaks at the same point. This restriction makes it difficult to analyze density estimation processes, and the linear reward-penalty scheme can legitimately been seen as a density estimation process, because during the estimation process, the peaks in the estimates tend to move around a bit. We do not start with knowledge of which of the actions is going to be the most likely, that would defeat the purpose of the learning algorithm, and if we incorrectly posit that one action is the most likely, and then later learn that another action is actually more likely, then the set of states that we move through is no longer directed.

The problem lies in the fact that the learning function that is used to update the states depend on the response signal, which is generated through a probabilistic process. Therefore, in general, the same action will not always generate the same response. Therefore even for a single event e, we cannot say anything about the relationship between $\mathbf{s}$ and $f_e(\mathbf{s})$. For example, if we pick action i where in our current state $\mathbf{s}$ we have $s_i = s^+$, and we observe the event $e = (i, 1)$, then $\mathbf{s} \sqsubseteq f_e(\mathbf{s})$ in both the Bayesian and implicative order. However, if we observe the event $e = (i, 0)$, which we will observe with probability $1 - \pi_i$, then it may be the case that $f_e(\mathbf{s}) \sqsubseteq s$, but it might also be the case that $\mathbf{s}$ and $f_e(s)$ are not even comparable in either order, because we are decreasing the value of s_i and increasing the value of all of the other s_j's, and so if we end up with $f_e(\mathbf{s})_j \geq f_e(\mathbf{s})_i$ for some $j \neq i$, then we have now changed the index of our maximal element, thus making the two states incomparable. Similarly, if we have $s_i \leq s^+$, then we cannot say anything about $\mathbf{s}$ and $f_e(\mathbf{s})$ for either of the possible response signals that we might see.

We had hoped that we would be able to prove something interesting, such as that f_e was Scott continuous, even if just for certain restricted environments, such as when $\pi \in \max(\Delta^n)$, but unfortunately this just simply is not true, since f_e is not necessarily monotone for any environment, except for one very contrived situation. The only case where either of the orders is actually interesting is when the environment is a maximal element, and we start with the initial state $\mathbf{s} = \perp$. This means that one action will always succeed, and all other actions will always fail, and when this is the case, we get something worth writing about. Since the environment is no longer stochastic, calling this a stochastic learning automata is a bit of a stretch, but this really is the only situation in which we have any interesting results, and so we will just go with it.

Our main result is that if $\pi = e_i \in \max(\Delta^n)$ for some $i \in \{1, \ldots, n\}$, then for all $s, t \in \{u \in \Delta^n : u \sqsubseteq \pi\}$ and for all $e \in E$, we have that $s \sqsubseteq t \implies f_e(s) \sqsubseteq f_e(t)$ and $s \sqsubseteq f_e(s)$ with $s = f_e(s) \iff s = \pi$ and so f_e is monotone and increasing. This result holds for both the Bayesian and implicative order. Also, for This means that for limited set of states and this restricted environment, $\mathrm{fix}(f_e) = \pi$, that is if we iterate f_e starting from any state $s \sqsubseteq \pi$, then we will end up converging to π in the limit. In fact, if we start from any state, we will converge to π in the limit, but we can only prove this using our orders on classical states if we start from a states that is less than π. The proofs of these claims are trivial, and follow straight from the definitions, and so we will not include them here for the sake of not insulting our readers.

4 Conclusion

In this paper, we looked at two different orders on the classical states, and analyzed the domain theoretic properties of the structure that emerged. The first order, the Bayesian order, induces an exact dcpo, which differs from a continuous dcpo in how approximations of elements are defined. This difference causes some of our domain theoretic notions, such as the interpolative property of the way below relation to break down in certain situations. The Bayesian order can induce a continuous dcpo if we remove the boundary states, those with 0 elements, from the classical states, but this removes very important states, such as the maximal elements, from the set, and so this is not a very reasonable restriction. The implicative order, which is a continuous form of logical implication, does in fact induce a continuous dcpo when applied to the classical states, and also has a very simple, intuitive interpretation. A state is more informative than another if it places a higher weight on its maximal element, and a lower weight on all other elements.

We also looked at learning system, the stochastic learning automata, equipped with a very simple learning scheme, and analyzed how the states change as the system learns, but we found that both of the orders were too restrictive to allow for much to be said about the progression of the states. The main problem is that the maximal element in two states must be the same in order for them to be comparable, and so any learning system that can place the highest belief on one element, but then shift the weights such that the highest belief is a different element produces states that are not comparable. If we add in randomness to the system, then essentially nothing can be asserted about the progression of the states, except in some very restricted situations.

It would be interesting to try to find a learning algorithm such that the states progress in a meaningful way in terms of the orders on classical states that we've looked at, but we were unable to do so. There are some major barriers, namely those mentioned in the previous paragraph, that could actually make finding such an algorithm an impossible task. There are certain domains, such as searching or sorting, in which algorithms tend to always progress directly towards the correct solution, and in fact in Martin (2004), a quantum search algorithm, Grover's algorithm, is studied and the algorithm is shown to progress monotonically in the implicative order towards the correct solution. It is likely that the stochastic manner in which learning algorithms proceed may make assigning an order to the progression of the states a fruitless endeavor.

Appendix

A The Bayesian order on Classical States

This section contains the theorems, lemmas, definitions, and propositions from Coecke and Martin (2002) pertaining to the Bayesian order on classical states. As they have been transcribed here, the numbers do not match those in the original paper. The titles of the sections are taken from the paper. We had previously prepared this section in order to summarize the 91 page paper into just a few pages for our own convenience, and so we felt it would be appropriate to include it in this document.

A.1 Two states and the parabola

Definition A.1. The *classical n-states* are

$$\Delta^n := \{x \in [0,1]^n : \sum_{i=1}^{n} x_i = 1\},$$

where $x = (x_1, \ldots, x_n)$ and $n \geq 1$.

A.2 A partial order on classical states

Definition A.2. Let $n \geq 2$. The projection which collapses the i^{th} outcome is the partial map $p_i : \Delta^{n+1} \rightharpoonup \Delta^n$ given by

$$p_i(x) = \frac{1}{1 - x_i}(x_1, \ldots, \widehat{x_i}, \ldots, x_{n+1})$$

for $1 \leq i \leq n+1$ and $0 \leq x_i < 1$. It is defined on $\mathrm{dom}(p_i) = \Delta^{n+1} \setminus \{e_i\}$.

Definition A.3. For $x, y \in \Delta^2$, we order classical two states by

$$(x_1, x_2) \sqsubseteq (y_1, y_2) \Leftrightarrow (y_1 \leq x_1 \leq 1/2) \text{ or } (1/2 \leq x_1 \leq y_1)$$

and for $n \geq 2$, for $x, y \in \Delta^{n+1}$ we define

$$x \sqsubseteq y \Leftrightarrow (\forall i)(x_i, y_i < 1 \implies p_i(x) \sqsubseteq p_i(y)),$$

where i ranges over the set $\{1, \ldots, n+1\}$.

Definition A.4. Let $n \geq 2$. For $x \in \Delta^n$, we set

$$x^+ := \max_{1 \leq i \leq n} x_i \quad \text{and} \quad x^- := \min_{1 \leq i \leq n} x_i.$$

We have $x^- \in [0, 1/n]$ and $x^+ \in [1/n, 1]$.

Proposition A.5. *Let $x, y \in \Delta^n$ and e_i be the pure states in Δ^n.*

(i) If $x \sqsubseteq y$, then there is an index i such that $x_i = x^+ \leq y^+ = y_i$.

(ii) For any i, $x_i = x^+$ if and only if $x \sqsubseteq e_i$.

(iii) If $x \sqsubseteq y$ and $x^+ = y^+$, then $x = y$.

Theorem A.6. *Δ^n is a partially ordered set for each $n \geq 2$. Its maximal elements are the pure states,*

$$\max(\Delta^n) = \{x \in \Delta^n : x^+ = 1\},$$

and its least element is the completely mixed state $\perp := (1/n, \ldots, 1/n)$.

Lemma A.7 (**Degeneration**). *If $x \sqsubseteq y$ in Δ^n, then*

$$(x_i = 0 \implies y_i = 0) \;\&\; (y_i = y_j > 0 \implies x_i = x_j)$$

for all $1 \leq i, j \leq n$.

Definition A.8. The standard projections $\pi_k : \Delta^n \to [0, 1]$ are $\pi_k(x) = x_k$ for $1 \leq k \leq n$.

Lemma A.9. *If (x_i) is an increasing sequence in Δ^n, then*

(i) There is an index k with $\pi_k(x_i) = x_i^-$ for all i.

(ii) There is an index k with $\pi_k(x_i) = x_i^+$ for all i.

Definition A.10. A subset S of a poset is *directed* if it is nonempty and

$$(\forall x, y \in S)(\exists z \in S) x, y \sqsubseteq z.$$

A directed-complete partial order, or *dcpo*, is a poset in which every directed subset has a supremum.

Proposition A.11. *The classical states Δ^n are a dcpo. In more detail,*

(i) If (x_i) is an increasing sequence, then

$$\bigsqcup_{i \geq 1} x_i = (\lim_{i \to \infty} \pi_1(x_i), \ldots, \lim_{i \to \infty} \pi_n(x_i)).$$

(ii) Every directed subset of Δ^n contains an increasing sequence with the same supremum.

Definition A.12. A map $f : D \to E$ between dcpo's is *Scott continuous* if it is *monotone*

$$x \sqsubseteq y \implies f(x) \sqsubseteq f(y)$$

and it *preserves directed suprema*:

$$f(\bigsqcup S) = \bigsqcup f(S)$$

for any directed set $S \subset D$.

Corollary A.13. *A monotone map $f : \Delta^n \to E$ into a dcpo E is Scott continuous iff for each increasing sequence (x_i) in Δ^n, $f(\bigsqcup x_i) = \bigsqcup f(x_i)$.*

Corollary A.14. *The map $s : \Delta^n \to [0, \infty)^*$ given by $s(x) = -\log x^+$ is Scott continuous. It has the following properties:*

(i) For all $x, y \in \Delta^n$, if $x \sqsubseteq y$ and $s(x) = s(y)$, then $x = y$.

(ii) For all $x \in \Delta^n$, we have $s(x) = 0$ iff $x \in \max(\Delta^n)$.

(iii) For all $x \in \Delta^n$, we have $s(x) = \log n$ iff $x = \perp$.

A.3 Symmetries for classical states

Definition A.15. The set of *permutations* or *symmetries*, $S(n)$ is the symmetric group

$$S(n) = \{\sigma | \sigma : \{1, \ldots, n\} \simeq \{1, \ldots, n\}\}$$

of bijections on the set $\{1, \ldots, n\}$. The composition of $x \in \Delta^n$ and $\sigma \in S(n)$ is written $x \cdot \sigma$.

Definition A.16. A state $x \in \Delta^n$ is *monotone* if $x_i \geq x_{i+1}$ for all $i < n$.

Lemma A.17. *For states $x, y \in \Delta^2$, we have $x \sqsubseteq y$ iff there is a $\sigma \in S(n)$ of $\{1, 2\}$ such that $x \cdot \sigma = (x^+, x^-)$, $y \cdot \sigma = (y^+, y^-)$, and $x^+ y^- \leq x^- y^+$.*

Theorem A.18. *For $x, y \in \Delta^n$, we have $x \sqsubseteq y$ iff there is a $\sigma \in S(n)$ of $\{1, \ldots, n\}$ such that $x \cdot \sigma$ and $y \cdot \sigma$ are monotone and*

$$(x \cdot \sigma)_i (y \cdot \sigma)_{i+1} \leq (x \cdot \sigma)_{i+1} (y \cdot \sigma)_i$$

for all i with $1 \leq i < n$.

Lemma A.19. *The map $x \mapsto x \cdot \sigma$ is an order isomorphism of Δ^n for each $\sigma \in S(n)$.*

Proposition A.20. *For $n \geq 1$, we have an order isomorphism*

$$\Delta^n \simeq \{x \in \Delta^{n+1} : \pi_i(x) = 0\},$$

for any of the standard projections $\pi_i : \Delta^{n+1} \to [0, 1]$ with $1 \leq i \leq n + 1$.

Definition A.21. The monotone classical states are denoted

$$\Lambda^n := \{x \in \Delta^n : (\forall i < n) x_i \geq x_{i+1}\}.$$

For any permutation σ

$$\Delta^n_\sigma := \{x \in \Delta^n : x \cdot \sigma \in \Lambda^n\}.$$

Lemma A.22. *For $x, y \in \Lambda^n$, $x \sqsubseteq y$ iff $(\forall 1 \leq i < n) x_i y_{i+1} \leq y_i x_{i+1}$.*

Proposition A.23. *Let $n \geq 2$.*

(i) For each permutation σ, Δ^n_σ is closed under directed suprema in Δ^n.

(ii) For an increasing sequence (x_i) in Δ^n, there is a $\sigma \in S(n)$ with $x_i \in \Delta^n_\sigma$ for all i.

(iii) The natural map $r : \Delta^n \to \Lambda^n$ is a Scott continuous retraction whose restriction to Δ^n_σ is an order isomorphism $\Delta^n_\sigma \simeq \Lambda^n$ for each $\sigma \in S(n)$.

Definition A.24. A function $f : \Delta^n \to E$ is *symmetric* if for all $\sigma \in S(n)$, we have $f(x \cdot \sigma) = f(x)$.

Lemma A.25. *Let E be a dcpo. Then*

(i) Every function $f : \Lambda^n \to E$ determines a unique symmetric extension $\hat{f} : \Delta^n \to E$ given by $\hat{f} = f \circ r$ where r is the natural retraction.

(ii) Monotonicity, strict monotonicity, and Scott continuity are inherited by $\hat{f}$ whenever they are possessed by f.

A.4 Approximation of classical states

Definition A.26. Let D be a dcpo. For $x, y \in D$, we write $x \ll y$ iff for all directed sets $S \subseteq D$,

$$y = \bigsqcup S \implies (\exists s \in S) x \sqsubseteq s.$$

The *approximations* of $x \in D$ are

$$\mathord{\downarrow\!\!\downarrow} x := \{y \in D : y \ll x\},$$

and D is called *exact* if $\mathord{\downarrow\!\!\downarrow} x$ is directed with supremum x for all $x \in D$.

A *continuous* dcpo is exact, and in that case, the "way below" relation and our notion of approximation are equivalent. In addition, the two notions coincide on maximal elements.

Lemma A.27. *Let D be a dcpo. For each $x \in D$, the set $\mathord{\downarrow\!\!\downarrow} x$ is directed with supremum x iff it contains a directed set with supremum x.*

Proposition A.28 (The mixing law). *If $x \sqsubseteq y$ in Δ^n, then*

$$x \sqsubseteq (1-p)x + py \sqsubseteq y$$

for all $p \in [0,1]$.

Remark A.29. A *path* from x to y in a space X is a continuous map $p : [0,1] \to X$ with $p(0) = x$ and $p(1) = y$. A *segment* of a path p is $p[a,b]$ for $b > a$. Any monotone path into Δ^n with its Euclidean topology is Scott continuous. For instance, by the mixing law (Prop A.28), the *straight line path* from x to y,

$$\pi_{xy}(t) = (1-t)x + ty$$

is Scott continuous iff $x \sqsubseteq y$.

Lemma A.30. *Let $x \sqsubseteq y$ with $x \in \Delta^n$ and $y \in \Lambda^n$. Then*

(i) If $y_i > 0$ for all i, then $x \in \Lambda^n$.

(ii) If $x \ll y$, then $x \in \Lambda^n$.

Proposition A.31. *Let $r : \Delta^n \to \Lambda^n$ be the natural retraction.*

(i) If $x, y \in \Delta^n_\sigma$ and $x \sqsubseteq y$, then $\pi_{xy}(t) \in \Delta^n_\sigma$ for all $t \in [0,1]$.

(ii) For $x, y \in \Delta^n$, we have $x \ll y$ iff

$$(\forall \sigma \in S(n))(y \in \Delta^n_\sigma \implies x \in \Delta^n_\sigma) \text{ and } (r(x) \ll r(y) \text{ in } \Lambda^n).$$

Theorem A.32. *The classical states Δ^n are exact.*

(i) For every $x \in \Delta^n$, $\pi_{\perp x}(t) \ll x$ for all $t < 1$.

(ii) The approximation relation $\ll$ is interpolative: If $x \ll y$ in Δ^n, then there is $z \in \Delta^n$ with $x \ll z \ll y$.

Lemma A.33 (Partiality). *For each $x \in \Delta^n$, the set $\mathord{\uparrow\!\!\uparrow} x$ is nonempty iff $x_i > 0$ for all i.*

Proposition A.34 (Approximation of pure states). *Let $n \geq 2$.*

(i) For all $x \in \Delta^n$, $x \ll e_i$ iff $x = \pi_{\perp e_i}(t)$ for some $t < 1$.

(ii) For all $x \in \Delta^n$, $\mathord{\uparrow\!\!\uparrow} x$ is an upper set (ie. $(\forall y, z) y \in \mathord{\uparrow\!\!\uparrow} x$ and $y \sqsubseteq z \implies z \in \mathord{\uparrow\!\!\uparrow} x$) iff it is empty, all of Δ^n, or contains a unique pure state.

Remark A.35. An approximation a of a pure state x defines a region $\mathord{\uparrow\!\!\uparrow} a$ of Δ^n known in domain theory as a *Scott open set.*

Definition A.36. A subset U of a dcpo D is *Scott open* if

- U is an *upper set*: $(\forall x \in U)(\forall y \in D) x \sqsubseteq y \implies y \in U$, and

- U is *inaccessible by directed suprema*: For any directed set $S \subseteq D$,

$$\bigsqcup S \in U \implies S \cap U \neq .$$

The collection of all Scott open subsets of D is σ_D.

Lemma A.37. *For all $x \in \Delta^n$, $\uparrow x$ is an upper set iff it is Scott open.*

Remark A.38. The relationship between approximation, partiality, and purity can now be summarized as follows:

(i) The partial elements are those $x \in \Delta^n$ with $\uparrow x \neq \emptyset$.

(ii) For a partial element $x \in \Delta^n$, $\uparrow x$ is Scott open iff $x = \pi_{\perp e_i}(t)$ for some i and some $t < 1$ iff ($x = \perp$ or x approximates a unique pure state).

Thus, the 'totality' of a pure state x is largely explained by the fact that $\uparrow a$ is Scott open whenever $a \ll x$.

Lemma A.39. *A subset $U \subseteq \Delta^n$ is Scott open iff*

- *Any monotone path from $x \in U$ to a pure state lies in U, and*
- *The line from $\perp$ to $x \in U$ has a segment contained in U,*

and for pure states x, there is an equivalence between 'approximation of x' and 'Scott open set containing x': Given any $a \ll x$, the set $\uparrow a$ is Scott open, while given any Scott open U with $x \in U$, we can (by exactness) find an approximation $a \in U$ of x with $x \in \uparrow a \subseteq U$.

A.5 Entropy, content, and partiality

Definition A.40. A Scott continuous map $\mu : D \to [0, \infty)^*$ on a dcpo is said to *measure the content* of $x \in D$ if

$$x \in U \implies (\exists \epsilon > 0) x \in \mu_\epsilon(x) \subseteq U,$$

whenever $U \in \sigma_D$ is Scott open and

$$\mu_\epsilon(x) := \{y \in D : y \sqsubseteq x \ \& \ |\mu x - \mu y| < \epsilon\}$$

are the elements ϵ close to x in content. The map μ *measures* X if it measures the content of each $x \in X$.

Definition A.41. A *measurement* is a Scott continuous map $\mu : D \to [0, \infty)^*$ on a dcpo that measures $\ker \mu := \{x \in D : \mu x = 0\}$.

Definition A.42. A Scott continuous map $\mu : D \to E$ between dcpo's is said to *measure the content of* $x \in D$ if

$$x \in U \implies (\exists \epsilon \in \sigma_E) x \in \mu_\epsilon(x) \subseteq U,$$

whenever $U \in \sigma_D$ is Scott open and

$$\mu_\epsilon(x) := \mu^{-1}(\epsilon) \cap \downarrow x$$

are the elements ϵ close to x in content. The map μ *measures* X if it measures the content of each $x \in X$.

Definition A.43. A *measurement* is a Scott continuous map $\mu : D \to E$ between dcpo's that measures $\ker \mu := \{x \in D : \mu x \in \max(E)\}$.

Proposition A.44. *Let $\mu : D \to E$ be a measurement and x an object that it measures.*

(i) If $\mu x \in \max(E)$, then $x \in \max(D)$.

(ii) If $\mu x = \bot$, then $x = \bot$, provided $\bot \in D$ exists.

(iii) If $y \sqsubseteq x$ and $\mu x = \mu y$, then $x = y$.

(iv) If $x_n \sqsubseteq x$ and (μx_n) is directed with supremum μx, then $\bigsqcup x_n = x$.

In addition, the composition of measurements is again a measurement.

Proposition A.45. *The natural retraction $r : \Delta^n \to \Lambda^n$ is a measurement.*

Lemma A.46. *Let $x \sqsubseteq y$ be monotone classical states in Δ^n. Then there is $k \in \{1, \dots, n\}$ such that*

(i) $(\forall i < k) x_i \leq y_i$, and

(ii) $(\forall i \geq k) x_i \geq y_i$.

Theorem A.47. *Let $\mu : \Delta^n \to [0, \infty)^*$ be the Shannon entropy on classical states*

$$\mu x = -\sum_{i=1}^{n} x_i \log x_i$$

where the logarithm is natural. Then μ is a measurement. In addition,

(i) For all $x, y \in \Delta^n$, if $x \sqsubseteq y$ and $\mu(x) = \mu(y)$, then $x = y$.

(ii) For all $x \in \Delta^n$, we have $\mu(x) = 0$ iff $x \in \max(\Delta^n)$.

(iii) For all $x \in \Delta^n$, we have $\mu(x) = \log n$ iff $x = \bot$.

References

B. Coecke and K. Martin. A partial order on classical and quantum states. Technical report, Oxford University Computing Laboratory, 2002.

A. Edalat. Domain theory in learning processes. *Electronic Notes in Theoretical Computer Science*, 1:18, 1995.

K. Martin. The measurement process in domain theory. In *Automata, Languages and Programming*, pages 116–126, 2000.

K. Martin. A continuous domain of classical states. Technical report, Oxford University Computing Laboratory, 2004.

K. Martin. Entropy as a fixed point. *Theoretical Computer Science*, 350(2):292–324, 2006.

J. Mashburn. An order model for infinite classical states. *unpublished*, 2006.

K.S. Narendra and M.A.L. Thathachar. *Learning automata: an introduction*. Prentice-Hall, Inc. Upper Saddle River, NJ, USA, 1989.

M.L. Tsetlin. *Automaton Theory and Modeling of Biological Systems*. New York: Academic, 1973.