| | |
|---|---|
| Title | Recursive Bayesian formulation of operant behavior : a framework |
| Sub Title | |
| Author | 久保田, 新(Kubota, Arata) |
| Publisher | 慶應義塾大学大学院社会学研究科 |
| Publication year | 1997 |
| Jtitle | 慶応義塾大学大学院社会学研究科紀要 : 社会学心理学教育学 (Studies in sociology, psychology and education). No.45 (1997. ) ,p.41- 59 |
| JaLC DOI | |
| Abstract | |
| Notes | シンポジューム : Pattern recognition in humans and animals |
| Genre | Departmental Bulletin Paper |
| URL | https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN00069957X-00000045-0041 |

# RECURSIVE BAYESIAN FORMULATION OF OPERANT BEHAVIOR —A FRAMEWORK

*Arata Kubota**

A computational framework for formulating operant behavior is proposed from the viewpoint of conditional probability. The framework regards an operant situation as a set of dynamic relations between recursive environments and recursive occasions in a behavioral stream. Utilizing Bayes' Theorem as a logical and quantitative tool, it is shown that the formulation gives relations resembling the matching law and melioration account, computational basis of reinforcement, and a possibility to simulate operant behavior on the basis of the formulation.

**Key words:** operant behavior, reinforcement, schedules of reinforcement, Premack's Principle, matching law, melioration, probability of behavior, conditional probability, Bayes' Theorem, prior probability, posterior probability, recursive environments, recursive occasions, simulation

Life can only be understood backwards, but must be lived forwards. —Kierkegaard
I've often seen a cat without a grin, but a grin without a cat! —Alice
Rose is a Rose is a Rose is a Rose··· —Stein

In 1990 I visited B. F. Skinner at his Cambridge home. It was a few months before he finally retired from his brilliant research career and the world. He had a leukemia which would cause us to say a sad farewell to him, and I had a computer simulation of operant behavior which would lead me into a deadly struggle for answers during the following several years.

The simulation had two conceptual, self-rotating *rings*: repertory and runtime as shown in Fig. 1, and nicknamed Two Ring Machine. The repertory ring had an *ad libitum* organization of behavioral repertoires and the runtime ring would have a current organization of behavior under several reinforcement schedules. Each segment of rings contained a behavioral topography or unit. The main ideas of the simulation were 1) a behavior unit (for example, eating) which had been deprived of its occasion (food) would accumulate in the runtime ring after it had come from the repertory ring constantly and repeatedly, and 2) a behavior unit (bar pressing), which gave the eating unit the

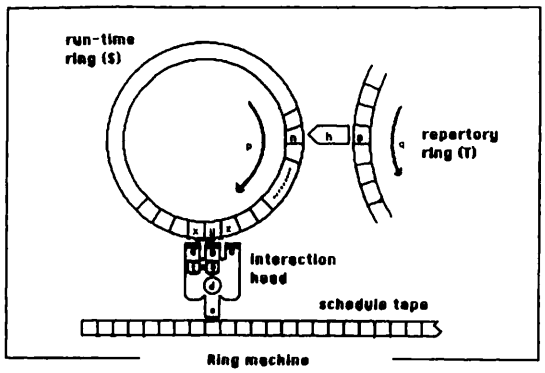



Fig. 1. Diagram of the simulation in 1990 (Two Ring Machine).

The simulation has two rings: repertory and runtime. The repertory ring had an *ad libitum* organization of behavioral repertoires and the runtime ring would have a current organization of behavior under several reinforcement schedules. An occasion-deprived behavior would accumulate in the runtime ring after it had come from the repertory ring, and an occasion-giving behavior would be copied and replace one of the accumulated behavioral units.

* Psychology, Fujita Health University, Toyoake, Aichi, Japan, e-mail to: akubota@fujita-hu.ac.jp, or akubota@amphora.iijnet.or.jp.

Some parts of this paper were presented in an evening session of the Jacksonville Conference on Behavior Dynamics, Jacksonville, Alabama, 1990, and Society for Quantitative Analyses of Behavior, Chicago, Illinois, 1997.

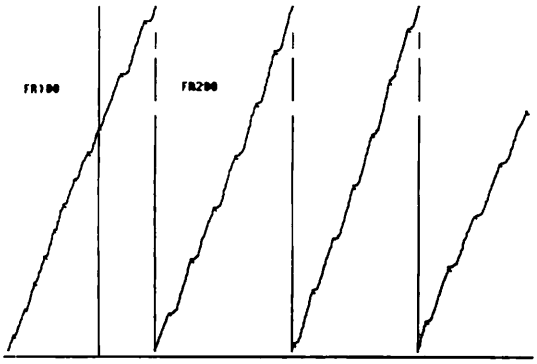



Fig. 2. Cumulative record generated by the simulation under FR100 and 200 schedules.

The post-reinforcement pause more clearly appeared under FR200.

occasion to occur, would be copied. The copy would be filled into a segment of one of the accumulated eating units that was adjacent and followed the occasion-giving behavior. Those were ideas hinted from Skinner's behavioral reservoir and Premack's Principle. The simulation was quite successful in generating characteristic cumulative records specific to several reinforcement schedules and to transient behavior seen during changes from one schedule to another. Some of the simulation results are shown in Fig. 2–4. For the results had I actually to devise some software tricks that had not been investigated and clarified fully in the behavior-analytical context, and some tricks were rather vague mathematically. The present research began then.

**About talking about probabilities**

People including psychologists frequently use a word *probability*, sometimes in a mathematically strict meaning and sometimes in a rather subjective, empirical manner. The exact definition of probability is usually hard to give. For instance, we talk about probability of behavior, but it is clear that there are several different kinds of probability for behavior. It depends mainly on the time point considered when we calculate probability and on the total probability space (denominator part in the calculation of probability) against which we locate the probability of the specified behavior. First, we may

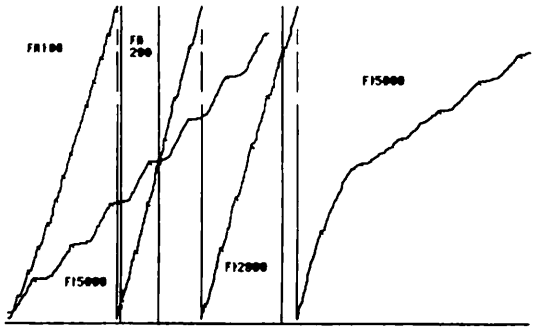



Fig. 3. Cumulative record at the transition from FR200 to FI2000 and 5000 schedules.

At the transition from FR200 to FI2000 no obvious change in responding patterns were observed, but after the schedule was changed to FI5000, a temporal decrease in response rate appeared and the scallop developed. Note that two traces are superimposed, and the unit of FI intervals was of a computational time.

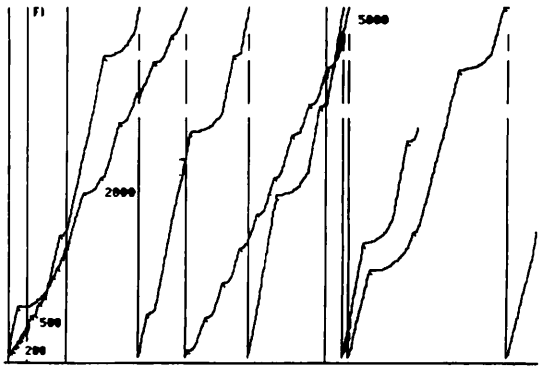
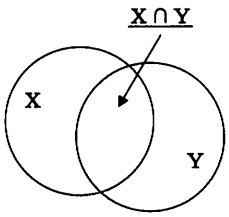



Fig. 4. Cumulative record under FI200, 500, 2000 and 5000 schedules.

The simulation showed typical scallops in some of the inter-reinforcement intervals under FI5000, in which clear scallops and short pauses appeared alternately.

calculate probability of behavior X after we have observed a behavioral series X, X, X, Y, Z, Z, Y, Y, X, X, and it will be 0.5 if we calculate it against the observed behavioral series in the above. Since the series has actually occurred already, the probability is firm ('firm probability' is a kind of oxymoron, though). But if we extend the calculation to unrealized future series, it will be either 0.5 or not. The latter case naturally includes a problem of estimation and




$$P(X|Y) = \frac{P(X \cap Y)}{P(Y)}, \quad P(Y)>0$$

Fig. 5. Definition of conditional probability.

prediction. Second, we may calculate the probability against time, but not against the behavioral units. Let X, Y, Z need 1, 2 and 3 second(s) respectively. The observed series will take 17 seconds totally, and the probability of X in the time series would be 0.29. Also, in this case we may try to extend the series to the future. In general, it seems reasonable to use the word *probability* as referring to prediction of future, unrealized events.

We encounter the above rather general problem also in interpreting and thoroughly quantifying Premack's Principle. Premack's Principle states that the probability of less probable behavior will increase if the behavior is accompanied or followed by more probable behavior. Consider how and when you estimate the probabilities of *less and more probable* behaviors. In general, they are what you obtain in the *ad libitum* situation, but when we try to extend the principle to the food-deprived situation as a possible reinforcement mechanism, a problem will appear: the probability of eating should be low in the food-deprived and even in the operant experimental sessions. Is the eating behavior more probable than the bar pressing behavior there? As for the observed behavior, the answer is 'the same or even less', and were it true, the occurrence of eating behavior could never reinforce the bar pressing behavior. We should state that the eating is more probable in the *ad libitum* situation sometime in the day but it is now *deprived and not yet realized*, or that the eating behavior is more probable if and only if its occasion (food) is available. In any case, the probability of more probable eating is not based on the data observed in the operant experimental sessions, and should be derived from somewhere else.

The present paper discusses those problems in more detail. Utilizing Bayes' Theorem as a convenient probabilistic tool, it proposes *recursive occasion theory* as a possible formulation of operant behavior and tries to formulate the reinforcement processes in some quantitative way.

**Conditional probability**

Conditional probability is defined as:

$$P(X|Y)=\frac{P(X\cap Y)}{P(Y)},\ P(Y)>0, \qquad (1.1)$$

where X and Y are events we would observe and $(X\cap Y)$ denotes a case that both X and Y occur simultaneously or contiguously, or anyway they occur as a set of the two events. $P(X\cap Y)$ is a joint probability of X and Y, and thus defined $P(X|Y)$ is a conditional probability of X when or after Y is observed (probability of X given Y). More thoroughly, conditional probability of X given Y is a probability with which the event X would occur under a condition that it is known the event Y has occurred. For convenience, hereafter let us write X,Y instead of $X\cap Y$ for the notation of logical conjunction (AND).

In relation to the above definition, there are two important rules (See Fig. 5):

$$P(X\cup Y)=P(X)+P(Y)-P(X\cap Y)$$
—additive law or sum rule (1.2)

$$P(X\cap Y)=P(X,Y)=P(X)P(Y|X)=P(Y)P(X|Y)$$
—multiplicative law or product rule (1.3)

If the events X and Y are independent of each other, $P(X,Y)=P(X)P(Y)$, and, using a conditional probability notation, this can be written as:

$$P(X)=P(X|Y) \text{ or } P(Y)=P(Y|X)$$
—independence between events (1.4)

Those relations can be immediately extended to a case of multiple events. For example, the product rule in a three-events case will be $P(X,Y|Z)=P(X|Y,Z)P(Y|Z)=P(Y|X,Z)P(X|Z)$.

One of the interesting behavioral implications of the concept of conditional probability becomes apparent if that concept is applied to the probability of eating when a food is available

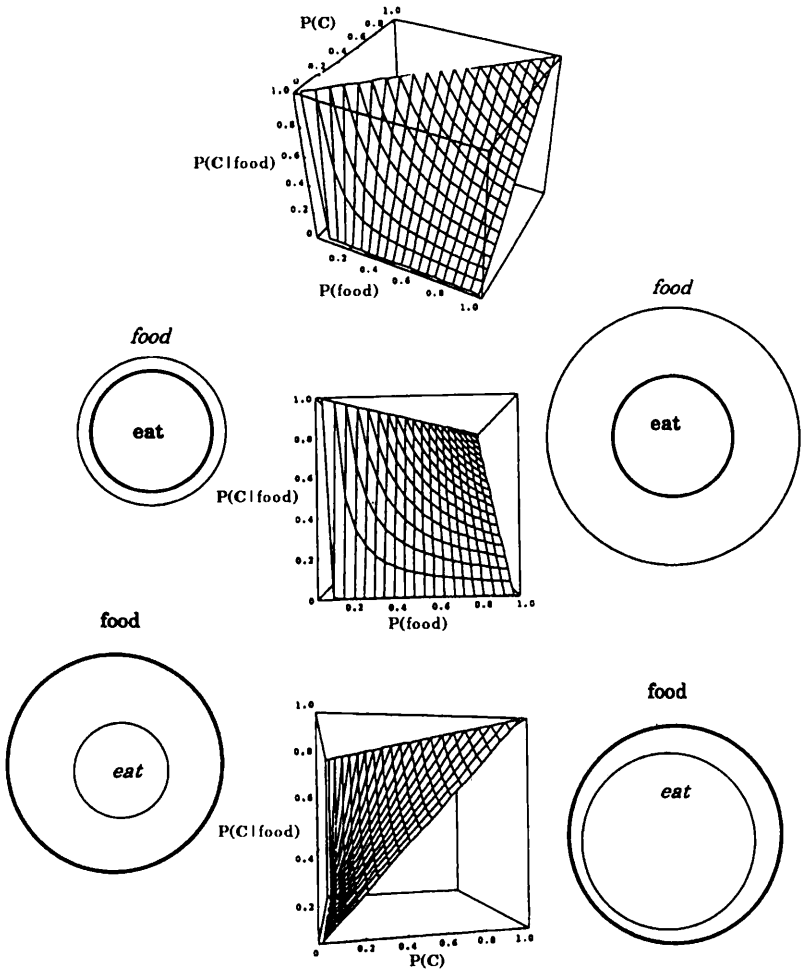



Fig. 6. Characteristics of P(eating|food).

In the middle row, P (eating) is fixed and P (food) varies. If the food availability gets lower, the P(eating|food) gets higher. If we allow P(eating) to get higher in this situation, we will get other hyperbolic curves that are nearer to the point (1,1). In the bottom row, P(food) is fixed and P(eating) varies. We get a simple, linear relationship between P(eating|food) and P(eating), and it will stand up from the diagonal line to the lines nearer to the vertical axis, when the food availability gets lower. The two aspects can be integrated into the top, three-dimensional plot. Note that P (eating|food) is indicated as P(C|food) in the plots.

(probability of eating the available food). Normally, eating is NOT independent of food availability and therefore P (eating|food)≠P (eating). In a usual and real situation, it is hard to think of the probability of *eating per se* when there is no food available, but here let us think of two probabilities P(eating) and P(food) as if we could estimate P(eating) separately from food availability. For example, P(eating) is higher in an animal that loves to eat than in an animal that does not like to eat. On the other hand, it is obvious that food availability varies from situation to situation. We can think then of the conditional probability of eating available food. Note some interesting relations between P (eating) and P (food).

In Fig. 6, two aspects of the conditional probability of eating available food are shown. The middle row shows the cases where P(eating) is fixed and P (food) varies. If the food availability gets higher, the P(eating|food) will be lower, and on the contrary, as the food availability gets lower, the P(eating|food) gets higher. Thus, we get a hyperbolic curve for one fixed value of P(eating). If we allow P(eating) to get higher in the above situation, we will get other hyperbolic curves that are nearer to the point (1, 1). Any hyperbolic curve formulates the probability with which eating behavior would occur along the food availability, deprivation/satiation continuum. In an environment with low food availability, an animal will immediately eat after it has found food. On the contrary, in an environment with high food availability, it is more likely that an animal will be engaged in the other activities without eating. The bottom

row shows the cases where P(food) is fixed and P (eating) varies. We get a simple, linear relationship between P(eating|food) and P(eating) for one fixed value of P(food), and it will stand up from the diagonal line to the lines nearer to the vertical axis, when the food availability gets lower. The two aspects are not independent of each other as we have seen, and they can be integrated into the top, three-dimensional plot.

Now is a good time to define two different *occasions* by means of conditional probability. Since we cannot eat without food, that is, P (eating) is fully included in P(food), the food is a *hard occasion* for eating. This case can be written as P(eating|food)=P(eating, food)/P(food)= P(eating)/P(food). See Venn diagrams in Fig.6. The other kind of occasion is a *soft occasion*, and when we let the occasion be an event O, P (eating) is not fully included in P(O). Sometimes the event O leads to eating, sometimes not, and sometimes eating can occur without the event O. See a Venn diagram of Fig. 5.

**Recursive environments**

Let us go back to our assumption that P(eating) can be estimated separately from food availability. As behavioral researchers, we would try to know P(eating) indirectly from P(eating|food) usually. Since P(eating) itself cannot be observed directly, it would be estimated by varying P(food) and observing P(eating|food). This is an *experimental* way of thinking. Premack's Principle does not depend mainly on this experimental thinking, but it basically depends on the ad libitum probability of eating, that is, it is based on an *observational* approach to behavior when food is always available or only softly deprived by setting on wheel-running some dependency for food availability. In a sense, the Premack's experiment, on which Premack's Principle is based, is a subtle and clever combination of experimental and observational ways of thinking.

Some difficulty will appear when we try to apply this principle directly to a fully experimental situation, in which we deprive animals of food as an explicit experimental operation. As mentioned before, P(eating) should be more probable than P(bar-pressing) if we expect that Premack's Principle can *verbatim* formulate the reinforcement of bar-pressing behavior after it has occasioned the presentation of deprived food and the occurrence of eating behavior. But it is never true unless we use the *ad libitum* probability of eating in the application of Premack's Principle to *experimental* situations, in which descriptive probability of eating is actually the same or less than the probability of bar pressing.

It is now obvious that it is necessary to overcome the gap if we like to have a logically and mathematically clear relationship between the 'more probable eating' in the ad libitum situation and 'less probable (less frequent) eating' in the experimental sessions. How can it be done at all?

The only way to the solution is to allow a few different probabilities of behavior and arrange them rationally in some sophisticated manner. Honestly, the secret solution is really simple and corresponds to what we usually do. We do not do it *sub rosa*, but we are not well aware of it. Let us think of two environments: an ad libitum environment (ALE) and an experimentally controlled environment (ECE). Also let us think of a few different probabilities of eating. One can be understood simply and it is an ad libitum probability (*Padlib*). This probability is based on the behavior observation and it is only descriptive and, by itself, never predictive for future occurrence of the behavior. Another is also descriptive, but describes the probability of the behavior that has actually occurred in the experimental sessions (*Pexp*). *Padlib* corresponds to ALE and *Pexp* to ECE. Note that both *Padlib* and *Pexp* are based on the observed data and fixed at the moment they are calculated, but the third probability has a different meaning. The third probability of behavior, *Ppotent*, is simply *Padlib* divided by *Pexp*:

$$Ppotent = Padlib / Pexp \qquad (2.1)$$

or in a logarithmic notation:

$$\log Ppotent = \log Padlib - \log Pexp \qquad (2.2)$$

In order to understand what *Ppotent* means, consider the case of ad libitum eating and food deprivation. In ALE, an animal can eat any time because food is always available. The probability of eating will simply conform to the P(eat-

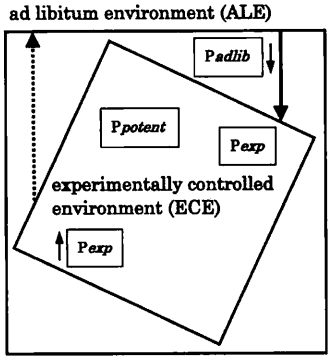



Fig. 7. Experimentally controlled environment (ECE) within ad libitum environment (ALE).

The occasion deprivation and schedule restrictions make ECE nested within ALE. As for the occasion-deprived behavior, *Pexp* < *Padlib*, and *Ppotent* is *Padlib* divided by *Pexp*. After a considerably long time period in experimental sessions, *Padlib* can be updated by taking *Pexp* into account, which reflects the consequences in the experimental sessions. Small arrows show the direction of parameter passing.

ing) and it is *Padlib* (eating). Then, if we deprive the animal of food, the hard occasion for eating, the animal may try to eat but cannot and, therefore, *Pexp*(eating) will be almost zero. The *Pexp* in an operant experiment is simply calculated on the basis of observed eating behavior that actually occurs when a food is presented. Usually this *Pexp* is also less than the *Padlib*. (Note the division would give a *Ppotent* value more than 1 in most cases. I will come back to this calculation method later.)

When *Padlib* is fixed, the less *Pexp* becomes, the more *Ppotent* would be, and the more *Pexp* becomes, the less *Ppotent* would be. Now we may notice that *Ppotent* is a kind of drive under situations where the access to food is restricted, and it is simply based on the two descriptive probabilities of behavior.

As *Pexp* is a result of some restricting operation added to *Padlib*, so the ECE is a result of the same operation added to the ALE. In other words,

$$Pexp = \text{a-restricting-function } (Padlib) \quad (2.3)$$

$$ECE = \text{a-restricting-function } (ALE) \quad (2.4)$$

From 2.1 and 2.3,

$$Ppotent = Padlib/Pexp = Padlib/\text{a-restricting-function } (Padlib). \quad (2.5)$$

Thus *Ppotent* becomes what reflects the restricting function well. If we assume that *Ppotent* is the higher probability of Premack's 'more probable behavior', the logical problem described above will be settled.

Since the ECE is generated through a restricting function of ALE, it is contained in the ALE diagrammatically (Fig. 7). But, since many aspects other than the restricting function are identical in both the ECE and ALE, ECE and ALE are similar to each other. The diagram with two squares can be extended to a diagram with more than two squares, that is, more than two nested environments. Let us come back to this later again. Because of the reasons to be mentioned then, a word *recursive* does not simply mean *repeating* in succession along time, but repeatedly *nested* as a structure.

$$P'(Ei \mid F) = \frac{P(Ei) \cdot P(F \mid Ei)}{\Sigma^{\infty}_{j=1} P(Ej) \cdot P(F \mid Ej)}$$

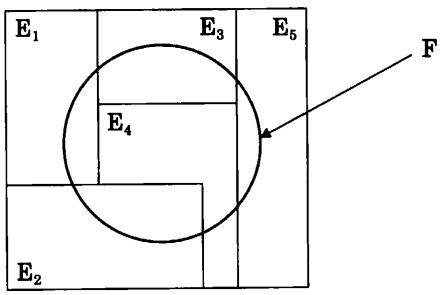



Fig. 8. Bayes' Theorem.

P ($E_i$) denotes an absolute or prior probability of the i-th event $E_i$, and $\Sigma^{\infty}_{j=1} P(E_j) = 1$. F is another kind of event, and $P(F \mid E_i)$ denotes probability of F given $E_i$, in other words, *likelihood* of F given $E_i$. $P(E_i \mid F)$ is called *posterior probability* of $E_i$ given F and sometimes called *probability of* $E_i$ *as a possible cause of* F. For convenience, probability as a possible cause is denoted with ' added: $P'(E_i \mid F)$. Note that $P'(E_i \mid F)$ is essentially a conditional probability though it has a few different names.

In the above relations between P*potent*, P*adlib* and P*exp* (2.1), I used a division operator without specifically mentioning why it should be division but not other operators, for example, subtraction, etc. It was simply based on the definition of conditional probability and for some convenience, and the idea that P*potent*, should be what reflects the difference or ratio between P*adlib* and P*exp*. Therefore it has not always been necessary to limit the operator to division up to now.

It is good now to notice that a hard occasion, for example food, would reverse the restricting function to a certain but limited degree. Thus, the hard occasion would make the ECE nearer to the ALE.

$$\text{occasions} = \text{a-restricting-function}^{-1} \quad (2.6)$$

The point is that, while we tend to think that an occasion is only what gives behavior a chance to occur in an environment, it actually changes the situation with two different environments and their relationship. Saying in a little bit exaggerated way, an occasion changes the world.

**Bayes' Theorem**

Bayes' Theorem or rule is simply another form of the definition of conditional probability in 1.1. From 1.1 and the product rule in 1.3, we obtain:

$$P(E_i|F) = \frac{P(E_i)P(F|E_i)}{\Sigma_{j=1}^{\infty} P(E_j)P(F|E_j)} \quad (3.1)$$

where $P(E_i)$ denotes an absolute or *prior probability* of the i-th event $E_i$, and $\Sigma_{j=1}^{\infty} P(E_j)=1$. All E's are mutually prime (independent). In this case $E_i$ is a discrete event and therefore the sum operator $\Sigma$ is used, and for a continuous event an integral operator $\int$ will be appropriate (Consider a case where the word *behavior* has no plural form). F is another kind of event, and $P(F|E_i)$ denotes probability of F given $E_i$, in other words, *likelihood* of F given $E_i$. See Fig. 8 for details of relations among F and $E_1$, $E_2 \cdots E_n$. From those we obtain $P(E_i|F)$, probability of $E_i$ given F. This $P(E_i|F)$ is called *posterior probability* of $E_i$ given F and sometimes called *probability of* $E_i$ *as a possible cause of* F. For clarity and convenience, probability as a possible cause is sometimes denoted with' added: $P'(E_i|F)$, which will also be used hereafter in the present paper. Note that $P'(E_i|F)$ is essentially a conditional probability though it has a few different names.

The event F in Fig. 8 can occur after $E_1 \sim E_5$. Probabilities of $E_1 \sim E_5$ may be different from each other and the likelihood of F given $E_1 \sim E_5$ is indicated as parts of $E_1 \sim E_5$ contained in the circle F (more in detail, those parts of $E_1 \sim E_5$ divided by $E_1 \sim E_5$). Those parts divided by F are *posterior probability* of $E_1 \sim E_5$ given F. In the sample situation of Fig. 8, $E_4$ is the main *cause* of F.

Bayes' Theorem is a convenient tool in Bayesian inference, where often $E_i$ is a hypothesis among several hypotheses (E's), its reliability or, sometimes, belief of a person who infers. It depends on the observation or realization of event F. Reliability of a hypothesis will be updated and increases each time an event which is favorable and supportive for the hypothesis has occurred and been observed. One posterior probability after a related event has been observed would be another prior probability in the next time. This is a frequently used method, and called Bayesian *updating*. It is natural to speculate that this updating can occur in the acquisition of a specified response in an operant experiment.

**Bayesian updating as a model of reinforcement mechanism**

In order to verify the speculation, let us consider the following situation. Let B be the behavior under consideration, for example, bar pressing, ~B (not B) be behavior other than B, and C be consummatory behavior or consequence. In order to assure that eating will occur immediately after a food presentation, P(eating|food) should be high, and to clearly show it, the bold **C** will be used to indicate a conjunction or joint event of food presentation and eating, that is, **C**=eating,food (or **C**=eating$\cap$food).

Several relationships in operant experimentation can be written in a form of conditional probability:

P(eating|food)+P(~eating|food)=1,
P(eating|food) gets higher after food deprivation, as shown already,

$P(food|B)=1$ if we set dependency on B for food (or in crf),

$P(food|B)>0$ and $P(food|\sim B)>0$ if we set contingency on B for food,

$P_n(eating|food)<P_{n-1}(eating|food)$ if some amount of food has been given,

and so on. The subscripts in the last relation, n and n−1, denote the steps along the time course. In most cases the subscript denotes the updating steps, n after an instance of reinforcement, n-1 before it, and so on.

Regarding that reinforcement is $P_n(B)>P_{n-1}(B)$, and that, as in the Bayesian updating, $P_n(B)=P'_{n-1}(B|\mathbf{C})$, the reinforcement process should include:

$$P_n(B)=P'_{n-1}(B|\mathbf{C})>P_{n-1}(B) \qquad (4.1)$$

Now let us seek a condition for making the relation in 4.1 true. From 4.1 and Bayes' Theorem,

$$P'(B|\mathbf{C})=\frac{P(B)P(\mathbf{C}|B)}{P(B)P(\mathbf{C}|B)+P(\sim B)P(\mathbf{C}|\sim B)} \qquad (4.2)$$

$$=P(B)\frac{P(\mathbf{C}|B)}{P(B)\,P(\mathbf{C}|B)+P(\sim B)P(\mathbf{C}|\sim B)} \qquad (4.3)$$

and therefore in order to make 4.1 true, $P(\mathbf{C}|B)>P(B)P(\mathbf{C}|B)+P(\sim B)P(\mathbf{C}|\sim B)$. Moving $P(B)P(\mathbf{C}|B)$ of the right hand to the left and using $P(\sim B)=1-P(B)$, we immediately obtain:

$$P(\mathbf{C}|B)>P(\mathbf{C}|\sim B), \qquad (4.4)$$

or

$$\frac{P(\mathbf{C}|B)}{P(\mathbf{C}|\sim B)}>1. \qquad (4.5)$$

This is the very condition we have sought for making $P_n(B)>P_{n-1}(B)$. What this condition means is self-evident. But if we add some explanation, this condition tells that, for one behavior being reinforced and its probability being increased, the probability with which that behavior occasioned eating,food joint event should be higher than that with which the other behavior occasioned the joint event. This rule should sound very reasonable to us behavioral researchers.

The relationship in 4.1 also applies to the behavior other than B, that is, ∼B. The ∼B version of 4.1 will be:

$$P'(\sim B|\mathbf{C})=\frac{P(\sim B)P(\mathbf{C}|\sim B)}{P(B)P(\mathbf{C}|B)+P(\sim B)P(\mathbf{C}|\sim B)}. \qquad (4.6)$$

and from 4.1 and 4.6, since the denominator parts of them are identical, we obtain:

$$\frac{P'_n(B|\mathbf{C})}{P'_n(\sim B|\mathbf{C})}=\frac{P_{n-1}(B)\,P_{n-1}(\mathbf{C}|B)}{P_{n-1}(\sim B)\,P_{n-1}(\mathbf{C}|\sim B)}. \qquad (4.7)$$

The left hand side of equation 4.7 is a ratio between new P(B) and P(∼B) as mentioned above, and it is a product of the old $P(B)/P(\sim B)$ and just updated $P(\mathbf{C}|B)/P(\mathbf{C}|\sim B)$ after C has occurred. P(B) and P(∼B) are *response rates* for behavior B and ∼B, and $P(\mathbf{C}|B)$ or $P(\mathbf{C}|\sim B)$ can be considered to represent *reinforcement rates* for behavior B and ∼B. More accurately, for some reasons that I have not yet mentioned, $P(\mathbf{C}|B)$ and $P(\mathbf{C}|\sim B)$ are not exactly the same as reinforcement rates, and therefore let us here call the ratio *likelihood ratio.* Equation 4.7 is our formulation of the reinforcement process at this moment.

Now let us compare it with a few of already established behavioral laws. First it is very similar to the Matching Law of the initial form:

$$\frac{B_1}{B_2}=\frac{C_1}{C_2}. \qquad (4.8)$$

But equation 4.7 includes a $P(B)/P(\sim B)$ ratio on the right hand side. Since the $P(B)/P(\sim B)$ ratio at one moment is the result of the former reinforcement processes, 4.7 can be re-written as:

$$\frac{P'_n(B|\mathbf{C})}{P'_n(\sim B|\mathbf{C})}=\frac{P_{n-1}(B)P_{n-1}(\mathbf{C}|B)}{P_{n-1}(\sim B)\,P_{n-1}(\mathbf{C}|\sim B)}$$

$$=\frac{P_{n-2}(B)\,P_{n-2}(\mathbf{C}|B)}{P_{n-2}(\sim B)\,P_{n-2}(\mathbf{C}|\sim B)}\cdot\frac{P_{n-1}(\mathbf{C}|B)}{P_{n-1}(\mathbf{C}|\sim B)}. \qquad (4.9)$$

Thus, the likelihood ratio of step n−2 and n-1 are included in the ratio of the current P(B) and P(∼B). It is obvious that equation 4.9 will be expanded to recursively include the likelihood ratio at steps n−3, n−4, and so on. Therefore, 4.9 will be:

$$\frac{P'_n(B|\mathbf{C})}{P'_n(\sim B|\mathbf{C})}=\frac{P_0(B)}{P_0(\sim B)}\cdot\frac{P_{n-1}(\mathbf{C}|B)}{P_{n-1}(\mathbf{C}|\sim B)}\cdot$$

$$\frac{P_{n-2}(\mathbf{C}|B)}{P_{n-2}(\mathbf{C}|\sim B)}\cdots\cdots\frac{P_0(\mathbf{C}|B)}{P_0(\mathbf{C}|\sim B)}, \qquad (4.10)$$

and in case that all the likelihood ratios of steps 0 to n-1 are identical (though this may not be an appropriate assumption when we think of the actual acquisition processes),

$$\frac{P'_n(B|C)}{P'_n(\sim B|C)} = \frac{P_0(B)}{P_0(\sim B)} \cdot \left(\frac{P(C|B)}{P(C|\sim B)}\right)^n. \quad (4.11)$$

To our surprise, this equation resembles the generalized form of the matching law:

$$\frac{B_1}{B_2} = K\left(\frac{C_1}{C_2}\right)^n. \quad (4.12)$$

Although I do not like to insist that our formulation 4.7 is identical to any of the two forms of the matching law at this moment, it is clear that our formulation can approximate them from a somewhat different viewpoint. The main difference is the time span with which we obtain and calculate the ratio of B and ~B. The Matching Law is an empirical law based on the behavioral data observed in a certain period, a session, for example. But our formulation has been derived rather theoretically and it concerns a more momentary time point, each time the reinforcement occurs.

If the main difference is of such a time, what can we say about on the relationship between our formulation and the Melioration account? The Melioration account tells that an organism allocates time to two or more response classes so that all local reinforcement rates are equal. Though it mainly concerns time allocation rather than response rates, the main logic of the Melioration account is a dynamic re-allocation of time for behavior, depending on the local reinforcement rate. In that sense, melioration occurs so that

$$\frac{\text{Time for } B_1}{\text{Time for } B_2} = \frac{\text{local } C_1}{\text{local } C_2}. \quad (4.13)$$

But, since it is a dynamic process, 4.13 may better be written as:

$$\frac{\text{Time for } B_1 \text{ at time n}/\text{Time for } B_1 \text{ at time n}-1}{\text{Time for } B_2 \text{ at time n}/\text{Time for } B_2 \text{ at time n}-1}$$

$$= \frac{\text{local } C_1 \text{ at time n}-1}{\text{local } C_2 \text{ at time n}-1}. \quad (4.14)$$

Immediately from 4.7, we obtain:

$$\frac{P_n(B)/P_{n-1}(B)}{P_n(\sim B)/P_{n-1}(\sim B)} = \frac{P_{n-1}(C|B)}{P_{n-1}(C|\sim B)}, \quad (4.15)$$

and therefore, if we assume that P(C|B) and P(C|~B) represent local reinforcement rates, and P(B) and P(~B) represent time allocations for B and ~B, equation 4.15 is almost identical to 4.14. At least, the dynamic aspect of the Melioration account also exists clearly in our formulation, and ours looks more precise in the time course than the Melioration account.

Starting from considering whether Bayesian updating can model reinforcement mechanisms and hence acquisition processes, we have found that our formulation has a deep and strong relationship to the Matching Law and Melioration account. Of course, as we have discussed, if P(C|B)>P(C|~B), then P(B) will increase, and it is suitable for an acquisition model. The Bayesian updating formulation of the reinforcement process is thus worth further study, but the above mentioned formulation is simply a prototype for constructing a model or formulation of more reality.

For example, the Bayesian updating formulation in 4.7 or 4.10 can be extended to multiple behavior situations, as the conditional probability and the single-step Bayes' Theorem can be easily extended to the multiple, successive events situations. That extension should have a close relation to behavioral chains and a possible formulation of the discriminative stimulus. There remain some subtle problems surrounding what P(B), P(C|B) and the power in equation 4.11 actually represent in real experimental situations. Further, those equations are merely a *logical* product in a sense, and we need some detailed, *mathematical/computational* framework in order to verify the logic in more realistic manners. I will discuss those problems in later sections, after we have integrated the recursive environments account and the above Bayesian updating mechanism.

**From single realization to distribution**

In the earlier section a non-descriptive probability P*potent* has been introduced in order to overcome the gap between observational and experimental approaches to the probability of behavior, that is, the problem we would encoun-

ter when we try to apply Premack's Principle to the food-deprived experimental settings. P*potent* was defined as P*adlib*/P*exp*. Since both P*adlib* and P*exp* are calculated on the basis of the observed, realized behavioral data up to the time point of calculation, they are determined to have a single value. This indicates that P*adlib* and P*exp* can be regarded also as realized and observed values. They seem to be fixed at a certain moment and to deterministically generate the probabilities on the next steps. A similar point can be made concerning Bayes' Theorem and the probability updating based on it. How about the non-descriptive P*potent*? Although it is a probability concerning future events and not yet realized, it seems to be fixed as a single value as if it were a realized value, because we obtain P*potent* simply with P*potent*=P*adlib*/P*exp*.

Is it really reasonable to think that those single values are the most appropriate ones to represent the real situations? Here is a very fundamental, probabilistic problem. If we observed a behavioral series X, X, X, Y, Z, Z, Y, Y, X, X, and we get 0.5 for P(X) based on the observed data, is it reasonable to think that the value 0.5 really represents P(X)? In our statistical convention, the above behavioral series should be regarded just as a sample realization of the probabilistic X, Y and Z. It is possible to obtain X, X, X, Y, Z, Z, Y, Y, X, Y in another observation and X, X, X, Y, Z, Z, Y, X X, X in the other. We need to estimate a distribution of X, Y and Z, from which each realization is drawn, as we usually do in most of the conventional statistical tests. Now is an appropriate time to step up from the single-valued probabilities of behavior to the probabilistic distributions of behavioral probabilities for a more realistic formulation of operant situations.

Each behavioral unit has two values, to occur and not to occur, and a behavioral series X, X, X, Y, Z, Z, Y, Y, X, X can be expressed as X, X, X, ~X, ~X, ~X, ~X, ~X, X, X from the viewpoint of one behavior X. Similarly, a behavioral series B, B, B, B, B, B, B, B, B occasions **C** sometimes in a manner of **C**, **C**, **C**, ~**C**, ~**C**, ~**C**, ~**C**, ~**C**, **C**, **C**. Therefore it is natural to regard that probability of X to occur x times in a total of n behavioral occasions, or probability of B to occasion **C** x

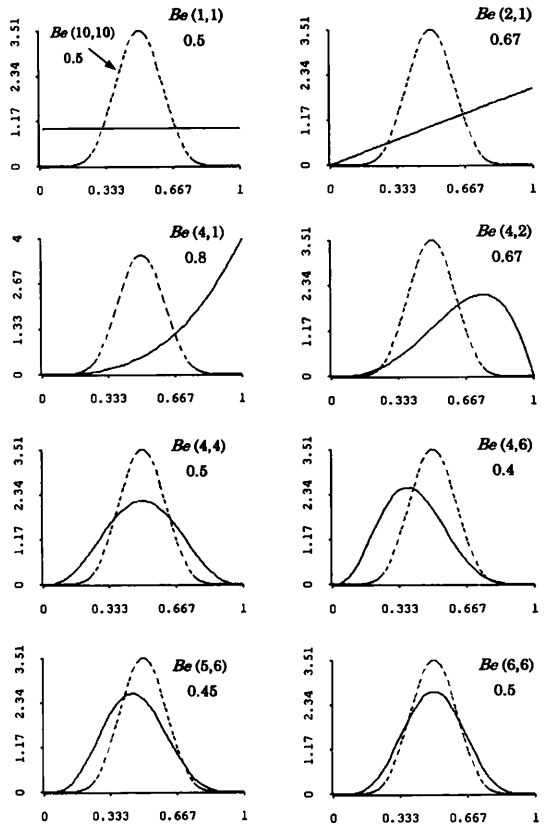



Fig. 9. Changes in shape and expectation of beta distribution for various sets of the two parameters.

Dotted curves in each graph represent *Be* (10, 10), and values of expectation of the distributions (average) are also shown below the tags of Be($\alpha$, $\beta$), which is shown in solid curves. In Be(1, 1) the shape of the distribution is flat, and it assures also high probability values. When a consequence favorable for the *hypothesis* comes out, the value of $\alpha$ increases and the distribution will be skewed to the higher probability direction (right). An unfavorable consequence increases the value of $\beta$ and hence transforms the distribution to the lower probability direction.

times in a total of n occurrences of B, conforms to a binomial distribution Bi(n, $\theta$). When $\theta$ is a parameter of the binomial distribution,

$$f(x|\theta)={}_nC_x\theta^x(1-\theta)^{n-x},\ x=0, 1, 2, \cdots, n. \quad (5.1)$$

The distribution version of Bayes' Theorem in

3.1 can be written as:

$$\omega'(\theta_i|z)=\frac{\omega(\theta_i)f(z|\theta_i)}{\Sigma_j\omega(\theta_j)f(z|\theta_j)}, \quad (5.2)$$

where $\omega(\theta)$ and $\omega'(\theta|z)$ are probabilistic distributions of prior and posterior probabilities of $\theta$, respectively, and $f(z|\theta)$ is a likelihood function. Since the denominator of 5.2 is evident, and regarding it is omissible, 5.2 can be written more simply as:

$$\omega'(\theta_i|z)\propto\omega(\theta_i)f(z|\theta), \quad (5.3)$$

and $f(z|\theta)$ transforms the prior probability distribution $\omega(\theta_i)$ to the posterior probability distribution $\omega'(\theta_i|z)$.

Since the likelihood function, which represents the probabilistic distribution of observed data, is a binomial function in 5.1, in order to select an appropriate prior and posterior distributions $\omega(\theta)$ and $\omega'(\theta|z)$, it is reasonable to look for them in a natural conjugate family of the binomial distribution. It is convenient and appropriate to select beta distribution for the prior and posterior distributions. Let them be beta distribution $Be(\alpha, \beta)$,

$$\omega(\theta)=Be(\alpha,\beta)=\theta^{\alpha-1}(1-\theta)^{\beta-1}/B(\alpha,\beta),\ 0<\theta<1, \quad (5.4)$$

where B(s, t) is a beta function and $B(s,t)=\int_1^0 u^{s-1}(1-u)^{t-1}du$. By 5.4, 5.3 will simply be:

$$\omega'(\theta|x)\propto\theta^{\alpha+x-1}(1-\theta)^{\beta+(n-x)-1}, \quad (5.5)$$

$$\omega'(\theta|x)=\theta^{\alpha+x-1}(1-\theta)^{\beta+(n-x)-1}/B(\alpha+x,\beta+(n-x)). \quad (5.6)$$

Since the posterior distribution $\omega'(\theta|x)$ is $Be(\alpha+x, \beta+(n-x))$, the updating simply transforms $\alpha$ to $\alpha+x$ and $\beta$ to $\beta+(n-x)$ in the same beta distribution.

For example, in the case of B, B, B, B, B, B, B, B, B occasioning **C** sometimes in a manner of **C**, **C**, **C**, ~**C**, ~**C**, ~**C**, ~**C**, ~**C**, **C**, **C**, the updating will start from Be(1,1) as the first prior probability, and the first occurrence of **C** converts it to Be(2, 1). In the same manner, it will be successively Be(3,1), Be(4,1), Be(4,2), Be(4,3), Be(4,4), Be (4,5), Be(4,6), Be(5,6) and Be(6,6). The distribution of the first prior probability has a flat shape, and it will change as shown in Fig. 9. Note that the expectation $E(\theta)$ of the beta distribution (average) is:

$$\mu=\alpha/(\alpha+\beta), \quad (5.7)$$

and it also changes while the updating of the distribution progresses.

This rather convenient updating process is realized by selecting a beta distribution for the distribution of the prior and posterior probabilities. There could be other, more theoretically appropriate distributions for it, but in the present paper let us utilize the beta distribution as the most appropriate example at this moment.

**Recursive environments with Bayesian updating**

We have seen the updating of a probabilistic distribution in the example of B occasioning **C** and the change in the shape of distribution, which reflects the observed behavioral data. The same operation can be applied to the probability of **C** itself. Note that **C** is a joint event of eating and food presentation. If food is not available at all as in the deprivation period, the beta distribution of **C** will be skewed to the left, lower probability, that is, **C** becomes unlikely to occur. This corresponds to the lower *Pexp*(**C**). If we have a flat Be(1,1) or beta distribution somewhat similar to a normal distribution in its shape, for example Be(100, 100), as a *Padlib* distribution, we will be able to obtain the distribution of *Ppotent* by dividing *Padlib* by *Pexp*, for example Be(50, 1). Also in this operation, the beta distribution is very convenient because its divisions by the same beta distribution simply become subtraction operations on the parameters $\alpha$ and $\beta$ of $Be(\alpha, \beta)$. In the above example, *Ppotent* will be Be (50, 99), and the distribution of *Ppotent* will be skewed to the right, higher probability. This means that eating is more probable if there is a food available.

The mechanical determination of *Ppotent* is now put in the center of nesting of the experimentally controlled environment (ECE) within the ad libitum environment (ALE). ECE is nested in ALE by means of really mathematical transformations. It is important to note that *Padlib* and *Pexp* are based on the observed, realized behavioral data in the real experiment and also in a possible simulation. In the simulation we can prepare one behavior, for example, eating, and examining whether there is food

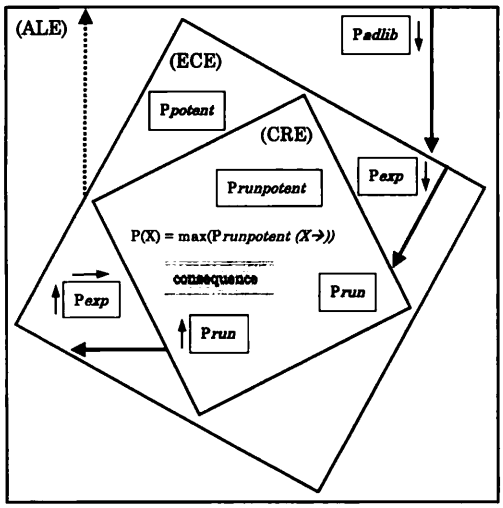



Fig. 10. Recursive occasion-setting environment (ROSE) diagram.

The recursive structure of the environments of ALE, ECE and CRE is shown by nested squares. Long arrows show directions of the nesting caused by occasion deprivation and schedule restrictions and of the returns from inner environments. Small arrows show the direction of parameter passing. Inheritance and update of the probabilities are almost identical to those in Fig. 7. The state of the behavioral system gets into ECE when some experimental controls are added, and gets back to ALE when the experimental controls have ended. Similarly, the behavioral system gets into CRE immediately after a deprived occasion is given, and gets back to ECE when another deprived occasion is given. The recursive occasion formulation assumes that the actual occurrence of behavior X conforms to the distribution of the highest *Prunpotent*(X→).

available or not, and if there is no food available, we add 1 to $\beta$ in Be($\alpha$, $\beta$) of *Pexp*(C), because eating cannot occur without food.

If we consider that the *Padlib* has been determined after a considerably long period of observation, it is reasonable to assume that *Padlib* will stay unchanged in the deprivation and experimental periods and that it is constant for a considerably long time period. While *Padlib* stays unchanged and if *Pexp* changes, then *Ppotent* will also change systemically. Of course, if deprivation or experimental sessions continue for a much longer time period than the period for which we have obtained *Padlib*, *Padlib* itself will change in accordance to the behavioral data obtained in the deprivation and experimental periods. Let us call this kind of change in *Padlib* an upward updating of *Padlib* by *Pexp*. This updating does not occur so frequently in usual situations.

As I have already mentioned in an earlier section, the most important and fascinating point of the recursive environments account is to allow another nesting of environment within the experimentally controlled environment (ECE). Tentatively, let us name the third environment within ECE the current runtime environment (CRE). CRE has *Prun* probabilities of behavior, as ECE has *Pexp*'s. *Prun* inherits a distribution of behavioral probability from *Pexp*, as *Pexp* does from *Padlib*. But, leaving *Pexp* unchanged, *Prun* will be updated precisely during an inter-reinforcement interval. The updating will occur after each occurrence of one behavioral unit, for example, B, and examination of its consequence, that is, for example, whether C occurs or not. Here also will be an upward updating of *Pexp* by *Prun* after a considerable session time. This upward updating is likely to occur more frequently than the upward updating of *Padlib* by *Pexp*. For example, it may occur after each occurrence of C, or after an experimental session has ended, etc. Note that an exact value for several 'considerably long time periods' has not been given.

It is highly important to point out that there is another probability *Prunpotent*, at the edge of the nesting of CRE within ECE. *Prunpotent* corresponds to *Ppotent* at the nesting of ECE within ALE. *Prunpotent* is calculated with *Prunpotent*=*Pexp*/*Prun*, and it will change precisely as *Prun* changes while *Pexp* and therefore also *Ppotent* stay unchanged for a considerably long time period. For example, *Prun* (eating, food | bar-pressing) decreases by each occurrence of bar pressing without getting food in an inter-reinforcement interval of an intermittent reinforcement schedule, and accordingly, *Prunpotent* (eating, food | bar-pressing) will gradually increase. This may have an important role in the non-homogeneous scattering of bar pressing

along the time course after an instance of reinforcement (Consider such a situation as an FI scallop).

For a summary of a slightly complicated set of nesting and upward relations, see Fig. 10. ALE contains ECE, and ECE contains CRE. The differences between ALE and ECE are the occasion-restricting functions or operations, which are, for example, food deprivation and other operant experimental settings including schedules of reinforcement. The difference between ECE and CRE is slightly subtle. Both ECE and CRE concerns experimental situations, but CRE concerns events in a short period such as an inter-reinforcement interval, and ECE concerns the history so far realized in a session or over sessions. The gap between ECE and CRE would clearly appear, for example, in the difference between overall reinforcement rate and local reinforcement rate, or difference between averaged reinforcement rate over several sessions and a current situation where an animal has not yet gained a food even after it has responded frequently. Anyway, CRE concerns the current, real-time system of behavior.

ALE has a set of *Padlib's* of many behavioral topographies or units. ECE has a set of *Pexp's*, which first inherit and correspond to *Padlib's* but are updated separately on the basis of the actual behavioral occurrences (and non-occurrences). In the same manner, CRE has a set of *Prun's*, which first inherit *Pexp's* but are updated separately on the basis of the current behavior stream. As *Padlib/Pexp* and *Pexp/Prun*, *Ppotent* and *Prunpotent* have been defined. *Ppotent* may represent a hunger level or such in an experimental session, and *Prunpotent* may represent more local changes in *Ppotent*. *Padlib* and *Pexp* will be updated *after a considerably long time period* upwards by *Pexp* and *Prun*, respectively.

Now the recursive structure of three nested environments should have become clearer. It is a simple interpolation of nesting to get a three-level nested structure from a two-level nested structure, even though the latter sounds more usual than the former. The three environments are almost isomorphic, and different slightly only in the functions connecting any two environments. This structure of recursive environments is the first milestone of our *recursive occasion theory.*

**What the reinforcement actually does**

Let us see what the reinforcement actually does in the framework mentioned so far. We have referred to a few changes accompanying the reinforcement. One of the changes is that a new probability of B becomes higher than the old probability of B. This has been expressed that a posterior probability of B, or probability of B as a possible cause of **C**, becomes higher than the prior probability of B. We have utilized Bayes' Theorem and the Baysian updating convention to formulate the relation between the prior and posterior probabilities of B. There, we have substituted P'(B|**C**) for P(B). Now let us examine this substitution more thoroughly.

See Fig. 8 again, in which let $E_4$ be B, and therefore P($E_4$) stands for P(B). Let F be **C**, and therefore P(F) stands for P(**C**). P'($E_4$|F) stands for P'(B|**C**). Accordingly, substituting P'(B|**C**) for P(B) is substituting P'($E_4$|F) for P'($E_4$). P'($E_4$|F) is actually a ratio of $E_4$,F and F, or simply $E_4$,F/F. Similarly, P($E_4$) is $E_4/(E_1+E_2+E_3+E_4)$. What does the substitution of $E_4$,F/F for $E_4/(E_1+E_2+E_3+E_4)$ mean? In a sense, it substitutes the contribution of a part of $E_4$ inside F for P($E_4$). Back to B and **C**, the updating substitutes the contribution of a part B inside **C** for P(B). This is neither that we substitute a part of $E_4$ inside F for $E_4$ itself nor that a part of B inside **C** for B itself. The substitution occurs beyond the level difference between P($E_4$) and P'($E_4$|F) and between P(B) and P'(B|**C**). The level difference means, in this case, that P($E_4$) and P'($E_4$|F) have different total probability spaces. Now we recognize that the substitution does not occur in a single dimension, but it may better consider that the substitution occurs from ad libitum context to **C**(or F)-oriented context. Is it just a baseless trick or a merely temporizing procedure for the formulation and a planned computer simulation? No.

Substituting P'(B|**C**) for P(B) is not simply to amplify P(B), but it is to temporarily re-organize the behavioral space (for example, $E_1+E_2+E_3+E_4$), which has been independent of **C**, to a new **C**-oriented behavioral space. It corresponds to the characteristic of our formulation that ECE is

nested within ALE by means of **C**-deprivation, and the food occasion reverses the nesting upwards. P'(B|**C**) can replace P(B) only within the **C**-oriented environment, that is, only within ECE (hence, also within ALE). Actually, this is the core concept on reinforcement in our formulation: the reinforcement does not occur within one environment but occurs beyond the border of environments.

This is not special only in our formulation. As described in the section on 'from single realization to distribution', a variable corresponding to P(B) is usually a *hypothesis* in the Bayesian convention, and usually it is not a mere probability of one event. It naturally includes some relations in it. Generally, substituting a probability of X as a cause of Y for the probability of X should mean that X has intrinsically some significance on Y. This is assumed in Bayesian conventions concerning belief or hypothesis verification.

In our case, P(B) starts as a simple probability of one event B. But, once it has been associated with **C**, P(B) is no longer a probability of such an isolated entity. B has become involved in a behavioral context or stream for **C**. Once B has been involved, yes, even once is enough, the updating of B is simply of a hypothesis that B occasions **C**, as in the Bayesian convention. The reinforcement process should be regarded as such a process. In a sense, reinforcement gives a new meaning to behavior B and its probability.

Let us call the behavior that has not yet been involved in a behavioral context or stream a *stream-free* behavior, and the behavior that has been involved a *stream-bound* behavior. The reinforcement process has two aspects: to make a *stream-free* behavior *stream-bound*, and to cause the changes in probability of the *stream-bound* behavior. As for the case of the hunger-eating-barpressing relation, some sensation related to hunger has a discriminative role and that discriminative property heads the **C**-bound behavior stream. When hunger goes away, the **C**-bound behavior stream will be decomposed.

Let us briefly examine the main logic of Premack's Principle in relation to the above core concept. Although we feel that both 'less probable behavior' and 'more probable behavior' belong to an identical single environment, it is not true. If they belong to an identical environment, we should be able to compare 'more probable behavior' and 'less probable behavior' directly, but it is impossible, as we have thoroughly discussed. More probable eating appears in a form of P*potent*, and bar pressing in a form of P*exp*. If we compare them directly, F and $E_4$ must appear in the same probability space competitively and they will never have an intersection or conjunction part in Venn diagrams like Fig. 8. We should consider that the reinforcement process naturally requires such epistemological conversion of the concept of behavioral probabilities.

Once we have admitted this core concept, we will be able to have a rather mechanistic reason why reinforcement can occur only in the situation where the more probable behavior accompanies less probable behavior as Premack has formulated. As mentioned above, the reinforcement process converts a stream-free behavioral space to a stream-bound (in this case, **C**-bound) one in our formulation. That is, the reinforcement process converts P(**C**) to the new total probability space and P(**C**,B) to the new P(B) as a part of the new total probability space:

$$\frac{P_n(B)}{1} \leftarrow \frac{P_{n-1}(\mathbf{C},B)}{P_{n-1}(\mathbf{C})} = P'_{n-1}(B|\mathbf{C}) > P_{n-1}(B). \tag{6.1}$$

If originally $P_{n-1}(B) > P_{n-1}(\mathbf{C})$ and if $P_{n-1}(B)$ includes $P_{n-1}(\mathbf{C})$ completely, then $P_n(B)$ will be 1 after the conversion. This is unusual in a behavioral system and impossible. When $P_{n-1}(B)$ does not include $P_{n-1}(\mathbf{C})$ completely, and if $P_{n-1}(\mathbf{C},B)$ is still higher than $P_{n-1}(\mathbf{C}, B_2)$, $P_{n-1}(\mathbf{C},B_3)$, etc., $P_n(B)$ becomes larger than $P_n(B_2)$ and $P_n(B_3)$, etc., and reinforcement will occur. But, after several reinforcements, P(B) will finally come to include P(**C**), and then $P_n(B)$ will again be 1. This is unusual in a behavioral system and is impossible. Only if $P_{n-1}(\mathbf{C}, B)$ is lower than $P_{n-1}(\mathbf{C},B_2)$, $P_{n-1}(\mathbf{C},B_3)$, etc., that is, reinforcement will not increase P(B), will it be assured that $P_n(B)$ will not be 1.

It is easy to think that $P_n(B)=1$ is unusual in a behavioral system, but is it still easy to regard it as impossible? It appears to be impossible only because of the requirements of our formulation. However, it is impossible not only in our formu-

lation but also in the real behavioral system. If $P(B)=1$, then probabilities of other behaviors, $P(B_2)$, $P(B_3)$, etc. will be 0. In that situation, $B_2$, $B_3$, etc. are completely deprived of occasions to occur. Now, $B_2$, $B_3$, etc. will become highly occasion-deprived and they become like **C**, and then $B_2$(or $B_3$, etc.)-bound stream will appear and P(B) will decrease. Generally $P_n(X)=1$ should and can not occur because of the reasons given above. Probably, due to the long evolutionary history of behavior, behavior systems may have come to reject a process that brings about a situation in which one behavior has a probability 1 (This may be related to the extinction process). P(B) should thus be less than P(**C**) originally, which will assure that a new P(B) falls in a normal range.

Standing on the above viewpoint on the reinforcement process, let us examine which of mere P(B) and P'(B|**C**) should be our main concern. That is actually the question of whether stream-free or stream-bound probabilities should mainly concern us. Consider a simulation of behavior. We prepare probabilities of behavior A, X, Y, Z, that is, P(A), $\cdots$ P(X), P(Y), P(Z). Usually we tend to think that P(X) or such probability changes in the behavioral system. We can think of another kind of probability like a probability with which behavior Z follows behavior Y, for example. P(B) at the beginning is a simple probability like P(X), but after P(B) has been involved once, that is, after it has become P'(B|**C**) in **C**-bound stream, it comes to look like a transition probability such as P(Y|Z), and so it is, by definition.

Since P(X) can be written as P(X|total-probability-space), conditional probability is more the general expression of probability. Therefore, a probability like P(Y|Z) may be more a general form than a simple probability, especially in systems like a behavioral system. Now we may consider that we should mainly concern ourselves with conditional probabilities like P'(B|**C**) rather than P(B), if P(B) can be expressed in a form like P'(B|**C**). As a notation it is easy, and we need to know whether it is possible as an expression of a distribution. Luckily, the beta distribution has a special shape when it is Be(1,1), that is, a flat distribution. The flat distribution with both parameters equal to 1 can be regarded to represent a probability like P(B), because it is the original distribution before any updates have occurred. Once behavior B is involved, parameters $\alpha$ and $\beta$ will be more than 1. Thus, it is possible and reasonable to think that a simple probability like P(B) can be expressed as the same distribution as P'(B|**C**).

We have paid a special attention to the apparently minor difference between P(B) and P'(B|**C**). This is not only because we consider that the difference is closely related to, and even the core of, our reinforcement process, but also because we have recursive environments (ALE and ECE) and they deeply concern the difference between ad libitum, simple probabilities and experimental stream-bound probabilities.

There are different transitions of one behavior X. For example, X→A, X→B, X→**C**, X→D, etc. We can think of P*adlib* (X→A), P*adlib*(X→B), etc. and also P*exp*(X→A), P*exp*(X→B), etc. Therefore, we can think of P*potent*(X→A), P*potent*(X→B), etc, also, and so forth for P*run* and P*runpotent*. While P*adlib* remains constant for a considerably long time period, P*exp* will be updated after reinforcements, etc., and therefore P*potent* will also change systematically. Now suppose that the behavioral system is **C**-bound, P*exp*(X→**C**) is low because the system is deprived of food and there are several restriction for **C** to occur due to reinforcement schedules, though it is already **C**-bound and not so low as in a simple deprivation period. There, P*potent* would be high, and in a very similar manner, P*runpotent* would be high, especially just before the reinforcement would occur.

There are different P*potent's* and P*runpotent's* of the transition from behavior X. For example, P*runpotent*(X→A), P*runpotent*(X→B), P*runpotent* (X→**C**), etc. Among them, if the behavioral system is **C**-bound, P*runpotent*(X→**C**) should be having the highest value.

Now let us make a seemingly bold assumption, which is actually the most important, key, assumption in our formulation. That is, *the occurrence of behavior X conforms to the probabilistic distribution of the highest Prunpotent*(X→). Here X→ denotes X→ (some behavior) and does not mean I made a mistake to drop A, B or such a character. The highest P*runpotent* (X→) is, in a **C**-bound behavioral system, P*runpotent*(X→**C**).

Although it might sound usual that the occurrence of behavior X conforms to P*runpotent* (→X), we take P*runpotent*(X→). It might sound strange, but really it is not, because, as we have thoroughly discussed earlier in this section, the reinforcement process tells us that we should substitute P*runpotent*(X→), and not P*runpotent* (→X), for Prunpotent(X). We have substituted P'(B|**C**) and not P(**C**|B), for P(B).

Several people say that the reinforcement process is a process of selection by consequence, as evolution is. This is a simple but beautiful statement, and I like to fully agree with it. But a statement that one behavior is selected by consequence is subtle and not so simple. Selecting one behavior by consequence has two aspects: making that behavior bound to a specific behavior stream and making that behavior more probable in that stream.

**Recursive occasion theory of operant behavior**

As for a recursion, there is a famous computational example: calculation of *factorial* of a natural number n. A computer program of the factorial function in a LISP dialect is:

```
(define factorial
    (lambda (n)
        (if (zero? n)
            1
            (* n (factorial (- n 1))))))
```

The algorithm is simple. When this factorial function is made to work with n being passed as an argument for it, as in a form of factorial(n), the function first examines if n is 0 or not. If n =0, then from definition of factorial it returns 1. If n is not zero, it would try to return a value of n times factorial (n−1), but the value factorial (n−1) has not been fixed, the function cannot return an exact value of n times factorial(n−1). So, the function calls itself downwards to know the value of factorial(n−1). Recursively called, the function first examined if n, which is actually n−1 this time, is 0 or not, and the same operations will be repeated until a recursively called factorial meets n=0. If it reaches n=0 then it returns 1.

This is an interesting example, and people tend to understand the algorithm intuitively when they come to the above 'If it reaches n=0 then return 1'. But the story has not yet ended. The returned 1 will be multiplied by 2, 3, ···, or n each time the called function goes back to the one-level upwards, and at the initial level, it returns the final value of factorial of n. The reason why this is possible is that each level of the function has kept a value passed as an argument to that level. The higher level passes a current value of the argument n when it calls itself downwards, it continues until n=0 is reached, and the function gets back step by step to the original level while it returns a current value of the answer.

Wise readers would now recognize that this is a leading metaphor for the recursive environments account. Each environment passes probabilities to its inner environment while it maintains their original values, and when a criterion is reached, each environment returns the result so far obtained to its outer environment. The criterion to be reached is, of course, that the highest P*runpotent*(X→) is realized as a concrete occurrence of X→ (some behavior). What are the operations like n-passing factorial(n) and result-returning n times factorial (n−1) in the behavioral world? The answer is now slightly clearer: they are to deprive and give occasions for behavior, and the reinforcement in which they act to calculate result probabilities in a Bayesian way and put them back.

The recursive nature of the nested environments should have been found also in the behavior stream, because it is organized in the interactions with the environments. In other words, the environment and behavioral system are mutually recursive. The key factor of the mutual recursion is an occasion, which mediates the environmental and behavioral, mutual recursions, and should be also recursive. Therefore it is suitable to call it a *recursive occasion*, the name of our formulation. The final part of the present paper is not to show results of the simulation based on the above formulation, but to touch recursive occasions in a behavior stream and their implications for understanding the behavioral chain.

From behavior-analytical knowledge accumulated so far, we know that there are several different kinds of behavioral chain. There is an explicitly controlled chain, and its example is a

chain schedule. Suppose we like to chain behavior A and B, and B occasions **C**, where an experimental dependency is set on A for B,**C** to occur, that is, unless A has occurred, B,**C** cannot follow. In this case, A is a 100% necessary condition for the occurrence of B,**C** (hard occasion). There are other cases where the dependency is less than 100%, where A occasions B,**C** with a higher probability, at least higher than the probability with which ~A occasions B,**C** (soft occasion). Probably we may use the term contingency for those cases. In a chain schedule the dependency is explicitly indicated by a stimulus change. There are also other cases where the dependency/contingency is not explicitly indicated by a stimulus change, as in a tandem schedule. A tandem schedule is not a chain schedule, but frequently B follows A and we regard that there is an implicit behavioral chain there. Accordingly, there are at least four types of behavioral chain: 1) B,**C** is dependent on A and it is explicitly indicated, 2) B,**C** is contingent on A and it is explicitly indicated, 3) B,**C** is dependent on A and it is not explicitly indicated, and 4) B,**C** is contingent on A and it is not explicitly indicated. It would be fine if we could formulate the four different cases in a uniform manner. Further, since a stimulus does not behave and it is clearly an environmental change, we may need to differentiate the cases with and without an explicit stimulus. Let us try.

Equation 4.2 consists of only B, ~B and **C**. Now, regarding the transition B→**C** as well established, let us extend 4.2 to include another behavior A. We obtain:

$$P'(A|B,C) = \frac{P(A)\,P(B,C|A)}{P(A)\,P(B,C|A)+P(\sim A)\,P(B,C|\sim A)}, \quad (7.1)$$

and we also obtain ~A version of 7.1. From the ~A version and 7.1, we obtain a ratio:

$$\frac{P'(A|B,C)}{P'(\sim A|B,C)} = \frac{P(A)\,P(B,C|A)}{P(\sim A)\,P(B,C|\sim A)} \quad (7.2)$$

This corresponds to 4.7, and in the same manner as in 4.7, the condition for P'(A|B,C) to increase is $P(B,C|A) > P(B,C|\sim A)$. We of course get the data for B,**C** from real experiments or simulation, and can calculate *Prunpotent*(A→B,**C**), to which A would conform. This extension includes a case in which the behavior A is actually B. In this case P(B→B) will get higher. Since B, **C** would almost always occur due to the high P (**C**), when we replace B for A and make 7.2 an equation on B→B, the ratio 7.2 would approximately get nearer to the ratio in 4.11, in which the original likelihood ratio is raised to a power. The power in the generalized matching law can also be understood in this direction.

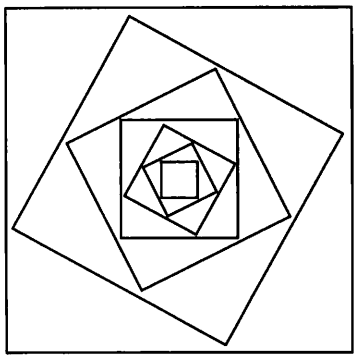


Fig. 11. ROSE diagram with more nested squares, at the core of which we meet the honey of reinforcement.

The ratio 7.2 has an interesting implication. Since it is the ratio between probabilities with which A and ~A would occur for B,**C**, if it is high(for example, more than 1), it shows the strength of the behavioral chain A→B,**C**. That is all right, but, if it is low (for example, less than 1), does that indicate the behavioral chain A→B, **C** has been weakened? We tend to think that a behavioral chain is first qualitatively connected and fixed, and then only its strength will change. But, the behavioral chain in our formulation is a probabilistic entity from the beginning to the end. If the probability for the chain A→B,**C** gets high, the behavior A simply overwhelms ~A. It is not true that two chains A→B,**C** and ~A→B, **C** exist solidly and one of them is selected. If the probability for the chain A→B,**C** gets lower than that of ~A→B,**C**, simply ~A takes over A in the single B,**C**-bound stream. There is no need to assume two chains as separate simulation entities. It looks as if there were a solid chain, but it only appears to be so.

The above extension does not include a case in which the dependency on A for B,**C** to occur is explicitly indicated by a stimulus change. An-

other extension to include such a stimulus is also possible. Since we have already used A, B, and **C**, let it be D, but to indicate that D is a stimulus, we write $S^D$. Similarly we obtain:

$$P'(S^D, A \mid B,\mathbf{C}) = \frac{P(S^D,A)\,P(B,\mathbf{C} \mid S^D,A)}{P(S^D,A)\,P(B,\mathbf{C} \mid S^D,A) + P(\sim S^D,A)\,P(B,\mathbf{C} \mid \sim S^D,A)}. \quad (7.3)$$

Note that the notation $\sim S^D,A$ denotes $\sim(S^D,A)$ and it includes two cases: $S^D$ exists but $\sim A$ occurs, and $S^D$ does not exist but A occurs. The joint event $S^D,A$ simply means that A occurs while $S^D$ exists, and that $P(S^D,A)$ increases means P(A) increases under $S^D$, because only P(A) could change while $S^D$ is a stimulus and $P(S^D)$ does not change for itself. As in 7.2, we can obtain a ratio $P'(S^D,A \mid B,\mathbf{C})/P'(\sim S^D,A \mid B,\mathbf{C})$, which would give a level of stimulus control by $S^D$.

Now we can write, in a LISP-like notation, the cases of 7.1 and 7.3 as:

(A occasions (B occasions **C**))

and

($S^D$ occasions (A occasions (B occasions **C**))).

Or boldly expanding **C**, based on the original meaning of it,

(A occasions (B occasions (food occasions eating)))

and

($S^D$ occasions (A occasions (B occasions (food occasions eating)))).

Since $S^D$ and food are occasions themselves, the meaning of the verb *occasion* may be slightly different from the case for the behavior A and B, but they are almost identical as an entity that gives an occasion for the following behavior.

The above extension for the behavioral chain and discriminative stimulus is just within the scope of our recursive occasion formulation of operant behavior, and no new devices need to be added for the actual simulations. The only additions we would do are new, smaller squares within the CRE, which correspond to those precise occasions mentioned above. The ROSE diagram of Fig. 10 will have more petals toward the center of a rose flower when we consider such smaller and recursive occasions (Fig. 11). At the core of a rose we will meet the honey of reinforcement and return.

Finally we should add some technical consideration on the selection of behavior that would occur actually in a simulation. The actual occurrence of behavior conforms to the distribution of *Prunpotent's*. There are many *Prunpotent's* in the system and they would compete for the final occurrence. If the distribution of *Prunpotent* of one behavior, say M, is skewed to the right, higher probability direction, the behavior would occur with higher probability. If the distribution of *Prunpotent* of another behavior, say N, is flat, and the expectation of the probability of M is higher than that of N, M would occur finally. But this is not always the case, because a flat distribution assures a high probability, too, and it can overwhelm the high probability from the distribution of the probability of M. A flat distribution assures a probability of almost 1, but a bell-shaped distribution like that of M does not easily give such an extremely high probability, even when it is skewed to the right. This is more likely when the skewness of the bell-shaped distribution gets smaller. Actually, this competition between skewed and flat distributions will contribute to the occurrences of non-reinforced behavior, as in the post-reinforcement pause. That is because the *Prunpotent* distributions of stream-free behavior are usually almost flat.

The above consideration implies that it is likely that we need not assume any temporal parameters in the present formulation of operant behavior. Rather, temporal factors are experimental free parameters we should verify in the simulation. Recall that we have not given an exact value for 'considerably long time periods' in the formulation so far mentioned. It is purposely reserved to be determined in the ongoing simulation work. Probably, time remains to be an eternal mystery, which we cannot easily simulate.

The main purpose of the present paper is to give a framework to formulate operant behavior and reinforcement process logically and quantitatively from the viewpoint of probability. Probabilities like *Prunpotent* are different from probabilities like *Pexp* and *Prun*. *Pexp* and

*Prun* accumulate the past data, that is, reinforcement history, but *Prunpotent* turns toward the future. The formulation of operant behavior and simulation based on it should grasp the future-oriented aspects of operant probabilities, which should be completely different from the respondent relations. Such probabilities might be in the simulation of 1990 in a form of accumulated eating units, but it was not so clear to me when I met Skinner. The years thereafter have been a time to struggle to answer the question that may have been asked by Skinner originally. He made a history and we live forwards.

**Acknowledgements**

I like to extend special thanks to the following people for the following reasons: To the late Professor B. F. Skinner for his warm words on my simulation attempts, to Dr. A. C. Catania for carefully reading the earlier, lengthy version of the present paper and detailed comments on it, to Dr. J. E. R. Staddon for having liked my computational puns, to Drs. E. G. Heinemann and S. Chase for improving many points of the manuscript text and making it readable, especially to the late Dr. D. A. Cook for his lasting friendship and discussions on the nature of operant behavior, and to many other people for their encouragement, all of which have been reinforcing my further challenge for the recursive occasion formulation. Rogo, ergo sum. (ak, 10/20/97)

**REFERENCES**

Catania, A. C. (1992) *Learning, the third edition.* Prentice Hall, Englewood Cliffs, New Jersey, USA.

Commons, M. L., Herrnstein, R. J., & Rachlin, H. (1982) *Matching and Maximizing Accounts, Quantitative Analysis of Behavior Series Vol. II.* Ballinger, Cambridge, Massachusetts, USA.

De Finetti, B. (1970) *Theory of Probability.* John Wiley & Sons, Chichester, England.

Ferster, C. B. & Skinner, B. F. (1957) *Schedules of Reinforcements.* Appleton-Century-Crofts, New York, USA.

Jaynes, E. T. (1995–1996) *Probability Theory: The Logic of Science.* Published in a form of internet WWW pages, Washington University. (http: //omega.albany.edu:8008/JaynesBook.html)

Premack, D. (1962) Reversibility of the reinforcement relation. *Science, 136,* 255–257.

Skinner, B. F (1969) *Contingencies of Reinforcements: A Theoretical Analysis.* Appleton-Century-Crofts, New York, USA.