Mexican Journal of Behavior Analysis
Special Issue on Operant Contingency
Vol. 21, 1995, pp. 163-274

Schedule Combinations and Choice: Experiment and Theory

Combinaciones de programas y elección: experimentos y teoría

J. E. R. Staddon[1]
Duke University, Durham, NC, USA

Abstract

This chapter has been about implicit and explicit choice. Implicit choice refers to the processes that determine the proportions of time that animals spend on different activities, the factors that maintain that distribution, and the effects of disturbing it by blocking activities or making access to one activity contingent on the performance of another. Explicit choice refers to special experimental procedures that pit two similar responses, such as pecking Left and pecking Right, against one another.

The first part of the chapter discussed temporal and stimulus control in the context of explicit choice between complex concurrent (choice) schedules. The first section showed how temporal control (in the form of *proportional* or *scalar* timing) and stimulus control combine in well-trained animals to produce the effect known as *conditioned reinforcement* on *chain* schedules. I also showed how conditioned reinforcers act as aids to memory when animals learn to respond on *delayed reinforcement* schedules and how memory limitations may underlie the effects of *second-order schedules*. Proportional timing seems to determine performance even on ratio schedules. I discussed in some detail a variety of experimental results on simple and *concurrent chain* procedures. The discussion showed that most, perhaps all, the concurrent effects do not represent choice in the usual sense at all. The animals do not seem to be comparing alternatives, but rather seem to treat each alternative as if it occurred in isolation. I was able to derive quite complex patterns of apparent preference and preference shift from an "ideal pigeon" who behaves according to proportional timing. I also showed how this analysis relates to the *optimal policy* on chain schedules, i.e., the pattern of responding that maximizes food rate: it turns out that proportional timing almost always produces a close-to-optimal pattern of

---

[1] Correspondence should be addressed to J.E.R. Staddon, Department of Psychology: Experimental, Duke University, Box 90086, 229 Soci/Psych Bldg., W9, Durham, North Carolina 27708-0086.

choice. I also showed the similarities between the optimal policy for animals on chain reinforcement schedules and the *optimal foraging theory* predictions about diet selection. The last part of this section discussed the *self-control* problem-preference for small-immediate vs. large-delayed rewards and showed how the same proportional-timing rule applies here also.

The last half of the chapter discussed implicit choice, the factors that determines the distribution of activities under free conditions. We saw that under many conditions the activity distribution is stable, and the organism resists in various ways perturbations that threaten to change the distribution from its *paired-baseline* level. The first attempt to understand these effects was made by David Premack, who concluded that higher-probability activities always reinforce lower-probability activities. This *molar* principle was extended first by the qualitative principle of *response deprivation* and then by a variety of quantitative optimality and economic analyses. The first of these, the *minimum-distance* model gave a special status to the paired-baseline levels or *bliss point.*

Optimality analysis is a general tool that can be applied to any adaptive system. It has allowed us to see common principles underlying implicit choice and explicit choice. Robust experimental findings such as the *matching law* turn out to be generally consistent with optimality models. Similar adaptive principles-diminishing marginal utility of reward frequency and amount-seem to underlie both the situations studied by Premack and more conventional schedules of operant reinforcement.

Despite their many successes, all optimality models fail under some conditions, because they are functional models, not models of mechanism. Animals and people are rarely, if ever, literal optimizers, systematically comparing the long-term payoffs associated with different policies. Thus, while matching on concurrent VI VI schedules fits in with a number of optimal policies, matching on concurrent VI VR does not. I described a number of other experiments in which animals clearly behave nonoptimally. The last part of the chapter therefore looked at the mechanisms of choice and behavioral allocation. The first conclusion was that *marginal* changes in *molar* variables probably do not have any direct effect on behavior, underlining the conclusion that even good optimality models, particularly molar optimality models, only describe what animals achieve, not how they achieve it. The last part of the chapter therefore focused on molecular mechanisms of behavioral allocation. I discussed three, *momentary maximizing*, *amelioration*, and *linear waiting*. The first and the last make very similar predictions in choice situations, but linear waiting promises to be more general. Quite apart from the quantitative details, it is clear that the expected time to the reinforcer, assessed through a memory-constrained timing mechanism, plays a dominant role in all the complex patterns of behavior generated by a variety of reinforcement schedules.

Key words: explicit choice, implicit choice, conditioned reinforcement, chain schedules, delayed reinforcement, behavioral allocation, behavioral economics.

Resumen

En este capítulo se analizó la elección implícita y explícita. La elección implícita se refiere al proceso que determina las proporciones de tiempo que los animales pasan en diferentes actividades, los factores que mantienen esa distribución y los efectos de distribuirla, bloqueando actividades o haciendo el acceso a una actividad contingente sobre la ejecución en otra. La elección explícita se refiere a procedimientos experimentales especiales que confrontan dos respuestas similares, tales como picar a la izquierda y picar a la derecha, una contra la otra.

En la primera parte del capítulo se discutió el control temporal y de estímulos en el contexto de la elección explícita entre programas concurrentes complejos (elección). En la primera sección se mostró como el control temporal (en la forma de *estimación del tiempo proporcional o escalar*) y el control de estímulos se combinan, en animales bien entrenados, para producir el efecto conocido como *reforzamiento condicionado* en los programas encadenados. También se mostró como los reforzadores condicionados actúan como facilitadores de la memoria, cuando los animales aprenden a responder en programas de *reforzamiento demorado* y como las limitaciones de memoria pueden subyacer a los efectos de los programas de *segundo orden*. Parece que la estimación del tiempo proporcional determina la ejecución aún en los programas de razón. Se discutió con algún detalle una variedad de resultados experimentales en procedimientos *encadenados simples y concurrentes*. La discusión mostró que la mayoría, quizás todos, los efectos concurrentes no representan elección en el sentido usual, en lo absoluto. Los animales no parecen estar comparando alternativas, sino más bien parecen tratar cada alternativa como si ocurriera en aislado. Se derivaron patrones bastante complejos de preferencia aparente y cambio de preferencia de un "pichón ideal", que se comporta de acuerdo a la estimación del tiempo proporcional. También se mostró como este análisis se relaciona con el *principio de optimización* en los programas encadenados, por ejemplo, el patrón de respuestas que maximiza la tasa de comida: resulta que la estimación del tiempo proporcional casi siempre produce un patrón de elección cercano al óptimo. Se mostraron las similitudes entre el principio de optimización, para los animales en los programas de reforzamiento encadenados y las predicciones de la *teoría de forrageo óptimo* acerca de la selección de dieta. En la última parte de esta sección se discutió el problema del *auto-control* —la preferencia por recompensas pequeñas pero inmediatas versus grandes pero demoradas— y se mostró como también se aplica aquí la misma regla de estimación del tiempo proporcional.

En la última mitad del capítulo se discutió la elección implícita, los factores que determinan la distribución de actividades bajo condiciones libres. Se vió que bajo muchas condiciones la distribución de actividad es estable y el organismo se resiste, de varias maneras, a las perturbaciones que tratan de cambiar la distribución de su nivel de *línea base apareada*. David Premack, hizo el primer intento por entender estos efectos y concluyó que las actividades con mayor pro babilidad

siempre refuerzan a las actividades con menor probabilidad. Este principio *molar* se extendió primero con el principio cualitativo de la privación de respuesta y después con una variedad de análisis cuantitativos de optimización y economía. El primero de estos, el modelo de *distancia mínima* dio un estatus especial a los niveles de línea base apareada o *punto de satisfacción*.

El análisis de optimización es una herramiento general que se puede aplicar a cualquier sistema adaptativo. Nos ha permitido ver principios comunes subyacentes a la elección implícita y explícita. Hallazgos experimentales robustos tales como la *ley de igualación* son generalmente consistentes con los modelos de optimización. Principios adaptativos similares —disminución de la utilidad marginal de la frecuencia y cantidad de la recompensa— parecen subyacer a las situaciones estudiadas por Premack y los programas más convencionales del reforzamiento operante.

A pesar de todos sus éxitos, todos los modelos de optimización fallan bajo algunas condiciones, porque son modelos funcionales, no modelos de mecanismos. Los animales y la gente rara vez, son literales optimizadores, sistemáticamente comparando los pagos a largo plazo asociados con diferentes principios. Así, mientras que la igualación en los programas concurrentes IV IV se ajusta con un número de principios de optimización, la igualación en los concurrentes IV RV no lo hace. Describí otros experimentos donde los animales claramente se comportan de una forma no óptima. En la última parte del capítulo, por lo tanto, se revisaron los mecanismos de la elección y la distribución conductual. La primera conclusión fue que los cambios *marginales* y las variables *molares* probablemente no tienen ningún efecto directo sobre la conducta, subrayando la conclusión de que aún los buenos modelos de optimización, particularmente los modelos de optimización molares, solo describen lo que los animales logran, no como lo logran. La última parte del capítulo, por lo tanto, se enfocó en los mecanismos moleculares de la distribución conductual. Discutí tres, la *maximización momentánea*, el *mejoramiento* y la *espera lineal*. El primero y el último hacen predicciones muy similares en las situaciones de elección, pero la espera lineal promete ser más general. Muy aparte de los detalles cuantitativos, es claro que el tiempo estimado para el reforzador, evaluado a través de un mecanismo de estimación del tiempo de memoria restringido, juega un papel dominante en todos los patrones complejos de conducta generados por una variedad de programas de reforzamiento.

Palabras clave: elección explícita, elección implícita, reforzamiento condicionado, programas encadenados, reforzamiento demorado, distribución conductual, economía conductual.

The study of choice is the study of the factors that make animals and people do one thing rather than another. Perhaps no topic in the field of operant learning has received more attention during the last ten years. In common speech, choice implies conscious deliberation. We often ponder decisions such as: Shall we go out for dinner or eat at home? Shall I stay home tonight and study or go out to a movie and have a good time? But animals also choose, and here the role of conscious comparison between alternatives is less clear. Whether the subjects are animals or people, the central scientific questions are (a) What measurable, external, factors determine choice? And (b) What mechanisms underlie choice behavior? These are the questions I will attempt to answer in the following pages.

I will look at choice in two kinds of situation, termed implicit and explicit choice. *Implicit* choice refers to the allocation of behavior under relatively free conditions. It is this kind of choice that is implied by real-life examples such as: How should the organism spend its time? How much time should it spend sleeping, eating, drinking, mating, and so on? Are there prefered patterns of time-allocation? Implicit choice is the study of the factors that determine the prefered pattern and indicate how it might be altered.

Explicit choice refers to special experimental arrangements that allow us to measure the effects on preference of factors such as reward amount, probability, delay and type. This chapter begins with an analysis of a variety of explicit choice procedures.

The chapter is divided into five parts. The first two parts are about explicit choice in chained schedules. They identify some simple general principles related to stimulus discrimination and reward delay. The third part is about implicit choice and the molar allocation of behavior. It concludes with a discussion of optimality models for the allocation of behavior under free conditions and on schedules of reinforcement. The fourth part takes up the optimality theme and shows how it is related to the economic approach to behavior. I discuss in this section the much-studied matching law and some potential explanations for it and related findings. It turns out that optimality and economic models can provide a comprehensive account for behavioral allocation although, like all functional accounts, they fail under some conditions. The last part of the chapter therefore takes up the question of mechanism, the moment-by-moment rules that animals use to adapt to reinforcement schedules. Recent advances are beginning to reveal what seem to be quite simple rules, including the reward-delay principle described in the first section, underlying the apparently complex behavior on reinforcement schedules. I conclude by pointing out some promising directions for future research.

Chain Schedules and Conditioned Reinforcement

In this section, I show how temporal discrimination and discrimination between nontemporal stimuli interact to determine behavior on procedures in which stimulus changes are produced by the animal's own behavior (chain schedules). Experimental results from some of these procedures have implications for theories of choice, human as well as animal, and for what has been termed self-control: refraining from choosing a small, immediate reward in favor of a larger, delayed reward. I begin by looking at how external signals modulate the effects of food delivery that is delayed or periodic.

Delay and Conditioned Reinforcement

Animals will use any available cue to help them allocate their behavior efficiently. For example, when food is delivered at fixed time intervals, animals behave differently during early and late parts of the interfood interval. During the late part of the interval they engage in food-related activities; during the early part they engage in activities less directly related to food. This is behavioral allocation that is under the control of time, measured from a time marker which, on FI schedules, is the delivery of food. Time acts like a stimulus that guides the allocation of behavior.

There is a similar pattern on schedules in which two or more exteroceptive stimuli, such as lights or sounds, alternate (multiple schedules). Food-related behavior is more likely in the presence of the stimulus that signals food; other activities occur in the presence of the stimulus that signals no-food.

Both these processes, control by time and control by a food-signalling stimulus, are involved in more complex procedures called chain schedules, in which stimulus changes are under the control of the animal. The complex patterns of behavior on chain schedules and related procedures illustrate four functions of stimuli:

(a) stimuli as aids or impediments to memory;
(b) stimuli as sources of value (conditioned reinforcement);
(c) stimuli as guides to behavioral allocation (behavioral contrast);
(d) stimuli as time markers.

These four functions, and the behavioral mechanisms associated with them, will allow us to make sense of a wide variety of experimental results from complex reinforcement schedules. I begin with chain schedules.

I will introduce chain schedules by first considering a simpler procedure called delayed reinforcement. Suppose we attempt to train a pigeon to peck a White key for food reward, but delay the food for 5 seconds after each peck. This is a *delayed-reinforcement* schedule. Even if the animal does eventually peck the key, the effect of each reward is likely to be small. The pigeon will make few pecks and will consequently get few food deliveries. The functional reason for the weak effect of delayed reinforcement is that it is difficult for the animal to pick out the peck, which is a brief event preceded and followed by other activities, as the best predictor of food that occurs several seconds later (in artificial-intelligence studies of learning mechanisms this is called the *assignment-of-credit* problem). Other events or behaviors, closer to the time of reinforcer delivery, are likely to be selected as predictors. But since the peck is the only effective response, the rate of reinforcement drops if pecking is displaced by other activities, so that the pigeon may eventually cease to respond actively at all (i.e., behavior extinguishes). How can we make the delayed-reinforcement schedule easier for the pigeon?

Chain Reinforcement Schedules:
Stimuli as Memory Aids and Sources of Value

We can make the bird's task much easier by altering the procedure slightly. Suppose that pecks on the White key are now immediately followed by a change in the color of the key to Green. After 5 seconds of a Green stimulus, food is presented as before. That is, a peck now produces a stimulus change plus food after 5 seconds. After food, the White key reappears (Figure 1). This procedure is called a *chain* schedule because the first contingency for food, a response, is linked to an additional contingency, in this case time, that is associated with a signal-the Green light. The time relations between pecking and food are exactly the same as in the delayed-reinforcement procedure yet the pigeon will rapidly learn to peck the White key if given 5 seconds of Green as a prefood signal.

Figure 1

FOR POSITION

Two-link chain schedule.

The change from White to the Green stimulus here is used as a reinforcer and seems to act like one. However, the Green stimulus gains its power not innately (or early in development), as does food reinforcement, but by virtue of its pairing with food. Therefore, a food-signalling stimulus is termed a *conditioned* or *secondary* reinforcer (to be contrasted with *primary* reinforcers, such as food). Why is this conditioned-reinforcement procedure more effective than the simple delay procedure?

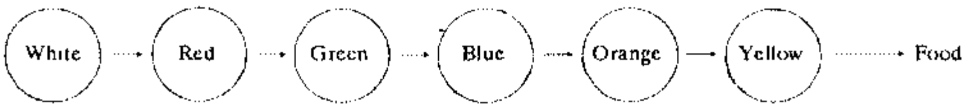
The Green-key conditioned reinforcer appears to aid conditioning for two reasons. First, it bridges the temporal gap between the peck (the real cause of the food) and its consequence (the food)-it is an aid to memory. Rather than having to remember a brief event occurring 5 seconds before its consequence, the animal has only to remember that pecking leads to stimulus change (and that the stimulus is associated with food). Because the peck-stimulus-change delay is negligible, this presents no difficulty. Second, because the rate of food delivery in the presence of the Green stimulus is much higher than in the presence of the White stimulus, the Green stimulus acquires a higher value than the White. The value is relatively transient and depends upon reliable delivery of the food. If food ceases to occur, the Green stimulus will quickly lose its value. The pigeon pecks to produce the Green stimulus because by doing so he turns off a stimulus with a low (zero) rate of food delivery and turns on one with a higher rate.

There are procedural tricks that will allow conditioned stimuli to retain their value for some time after primary reinforcement has ceased. These tricks depend upon memory confusions like those discussed in a moment in connection with second-order schedules (See discussions of *second-order conditioning-for example*, Rescorla, 1982).

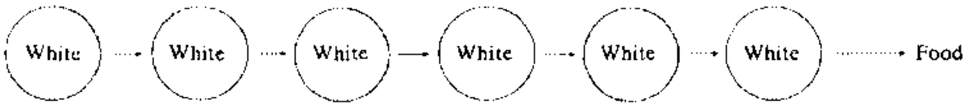
The factors that maintain behavior are often very different from those that give it birth. Our pigeon learns to peck for the Green stimulus because it has more value than the White stimulus it replaces. But once the bird has learned to peck the key, other factors come in to play. The duration of the Green stimulus is fixed, for example. Thus time can begin to play a role, and the role it plays is in some respects surprising. What would you expect to be the effect of increasing the duration of the Green stimulus from 5 to 10 s, for example? Will this change have any effect on how long the animal waits in the presence of the White key before pecking to produce Green? It makes no sense for the animal to wait for any but the shortest time in White, since the delay-to- food in Green is fixed; every second the bird waits in White unnecessarily delays food by the same amount. Nevertheless, the longer the Green stimulus, the longer the bird will wait in White before pecking to produce it (Wynne & Staddon, 1988). This is an example of a mechanism called proportional or scalar timing that plays a dominant role in chain-schedule performance. I describe proportional timing in a moment. Now let's look in more detail at the factors that maintain (as opposed to initiate) performance on chain schedules. We will see that the function of stimuli as sources of value is much less important than other functions, such as stimuli as memory aids or impediments and stimuli as time markers.

Pigeons on multilink chain schedules behave in ways that are not fully explained by the idea that pecking in link N is reinforced by the production of link N+1. Consider a chain schedule with three links: pecking in link 3 (S_3) is reinforced by food; pecking in link two (S_2) is reinforced by the production of link three; pecking in link 1 (S_1) is reinforced by the production of link 2. There is nothing in principle wrong with extending such a chain indefinitely, to six links, for example (Figure 2, top). Suppose that each link (S_1, S_2, etc.) is T-seconds long; i.e., the first peck after T-s in the presence of S_1 causes it to change to S_2, and so on-this is called a chain FI schedule. How many such fixed-interval links can be strung together and still maintain responding in S1, the stimulus most remote from food?

Figure 2



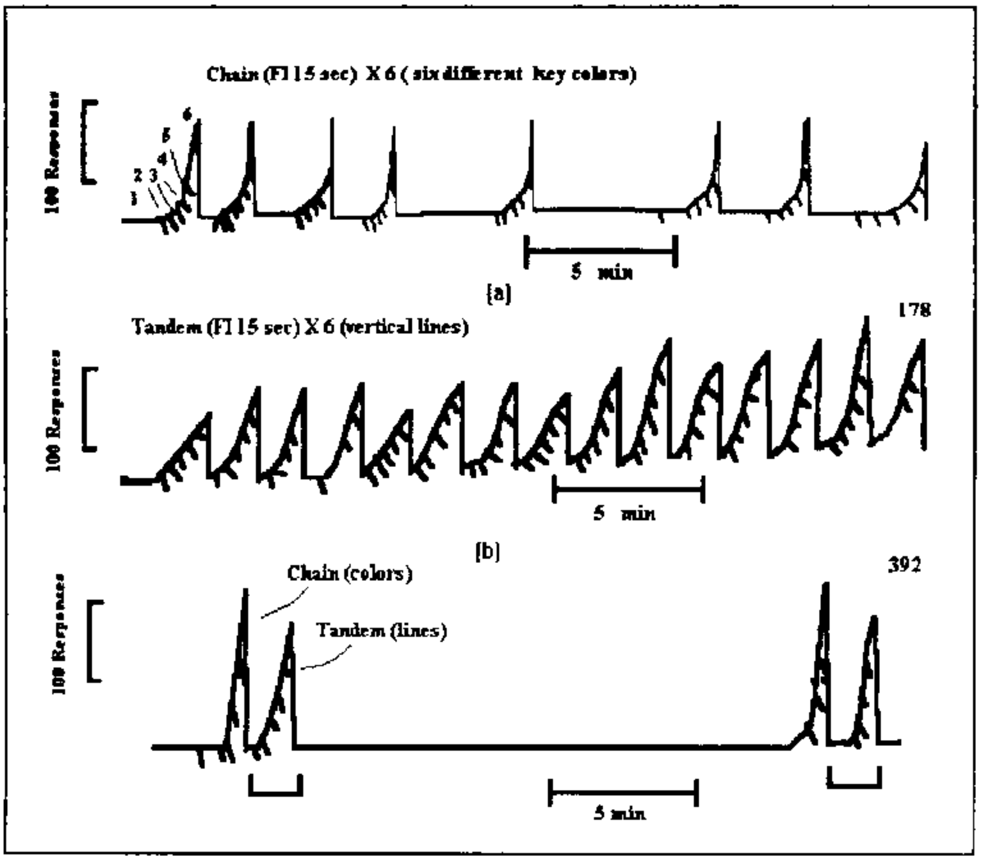
CHAINED SCHEDULE



TANDEM SCHEDULE

Six links of a FI 15 s schedule. Top: A 6-link chain; bottom: a comparable 6-link tandem schedule.

Figure 3



Top: Cumulative record from a well-trained pigeon on a 5-link chain FI 15 s schedule. The record resets after food delivery. Middle record: Performance on a comparable tandem schedule. Bottom: Record from a pigeon exposed to both the chain and the tandem schedule, each signaled by appropriate stimuli (from Catania, 1970).

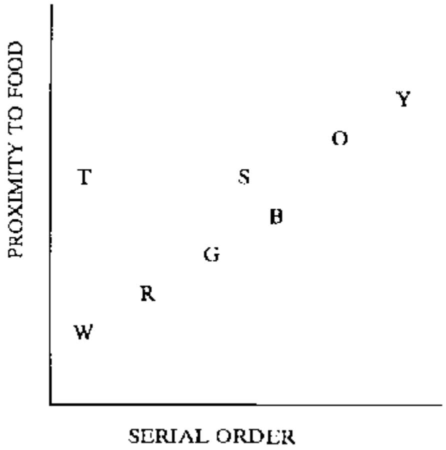
The answer, for even quite short T-values, is not more than five or six. The top record in Figure 3 shows a cumulative record from a pigeon trained with six fixed-interval 15-second links (there is a blip in the record to indicate each stimulus change; the recorder pen reset after food). Long pauses occur after food and the times between food delivery are always much longer than the 90-second minimum prescribed by the schedule. An additional link would have caused the pigeon to stop responding altogether. With longer fixed intervals, five links are the upper limit. What causes this breakdown? We can get some clues by changing the procedure slightly-by eliminating the different stimuli that signal each link in the chain. This new procedure is called a *tandem* FI schedule (Figure 2, bottom).

*Tandem Schedules. The middle record in Figure 3 shows typical perfor*mance on a six-link tandem FI schedule. On a tandem FI schedule, the timer for the second link doesn't start until the animal has responded after time T in the first link, and so on, but the stimulus on the response key is always the same. Performance on the tandem schedule is different from performance on the chain in two main ways. First, there are no pauses at the onset of each link-because there are no stimulus changes to serve as time markers. Second, the response rate at the end of the interfood interval is much lower on the tandem, but the postfood pause is much shorter than on the chain. You can see these differences in the cumulative records in Figure 3: the last fixed interval before food has a much steeper slope in the chain than the tandem record, for example. These differences appear even more clearly in the bottom record, which shows both types of schedule successively in the same pigeon.

Clearly, one problem with the simple conditioned-reinforcement idea is that it cannot explain results like those in Figure 3. Why should five or six chain links be the limit? Why should the pigeons respond *less* in the first chain link than in the first tandem link? These problems are usually handled by factors in addition to the value-added property of conditioned reinforcement. Three differences between chain and tandem schedules have been suggested as reasons for the very different behavior they produce. First, the relative proximity to reinforcement (food) of each stimulus in the series (stimuli early in the chain are far from the next food, later stimuli are successively closer). Second, the pairing, or lack of pairing, of a stimulus with food. And third, the contingency between pecking and a transition to a stimulus closer to food.

The last two factors, pairing with food and contingency, define conditioned reinforcement in the strict sense. Pairing with food is important because a reinforcer must have value, and value is derived from direct association with food. The contingent relation between response and reinforcer is important because reinforcers are usually assumed to act through response contingency. It is well known that once the response has been acquired, contingency is not all-powerful even for food reinforcement (animals well-trained to respond on a food-contingent schedule will persist almost indefinitely if the pattern of food deliveries is maintained, even if the response contingency is abolished, cf. Herrnstein, 1966), so we should not be surprised to see that contingency turns out to be relatively unimportant for conditioned reinforcement also. The pairing with food is important, as we will see in a moment.

Figure 4



Stimuli arranged according to their serial order (x-axis) and temporal proximity to food (y-axis) for three different serial procedures: chain schedule (color symbols), tandem schedule (T), and a scrambled chain schedule (S).

But the first factor, relative proximity of each stimulus to food, seems to exert the major effect (cf. Mazur, Snyderman & Coe, 1985; Shull & Spear, 1987). This is illustrated in Figure 4, which plots the serial order of each stimulus against its temporal proximity to food. The color symbols (see Figure 2) in the figure denote stimuli in the chain schedule. For this case serial order is perfectly correlated with proximity to food, and response rate is perfectly correlated with both: the closer to food, the more responding to the stimulus (see Figure 3). The "T" denotes the serial order and average proximity to food of the single tandem stimulus. Here the stimulus has an intermediate temporal proximity to food, and it sustains an intermediate rate of pecking. The "S", in the figure, represents the proximity to food of all six chain stimuli when their order is "scrambled" from interval to interval so that each stimulus occurs in each serial position 1/6 of the time. Once again, proximity to food seems to be the major factor, because all six stimuli sustain the same intermediate response rate under these conditions.

Catania (1979) has demonstrated the relatively weak role of the response contingency on chain FI schedules. So long as the response contingency for food is maintained in the final link of a chain FI schedule, the contingency can be omitted in the earlier links with little effect on key pecking. That is, the fixed-interval schedules can be replaced with a *fixed-time* schedule in all but the final link and the pigeon will continue to peck in each link much as before. If the stimuli truly serve as "reinforcers", in the sense that their effect depends on the contingent relation between peck and stimulus change, response rate should have decreased when the contingency was removed. It changed little, suggesting that the other factors, temporal proximity to food, and pairing with food, are more important.

Thus, data from chain and tandem schedules all point to relative temporal proximity to food as the major factor that maintains pigeon behavior. I will now show in more detail how this timing process works.

Dynamics of Chain and Tandem Schedules: Stimuli as Time Markers

All reinforcement schedules present stimuli and reinforcers to the subject in a temporal sequence. On most schedules this sequence is quite regular. We know that mammals and birds readily detect temporal regularity-the phenomenon is termed *temporal discrimination*. For example, we know that on FI schedules animals always pause before beginning to respond, and that the pause is roughly proportional to the interfood interval. We will term this temporal discrimination process *proportional timing*. The terms *scalar timing* (Gibbon, 1977) or *linear waiting* (Wynne & Staddon, 1988) have also been used for more sophisticated versions of the same idea.

Let's see how proportional timing might work on chain schedules. On a two-link chain FI T s, the first response after T s in the presence of S_1, the first stimulus after food, causes a change to S_2. The first response after T s in S_2 yields food. We know also that the pause is determined by time since a *time marker* which, on FI schedules, is just the delivery of the reinforcer. But on our two-link chain schedule there are *two* time markers, the reinforcer (usually food)-which determines the first pause-and then the transition from S_1 to S_2-which determines the second pause, in the final link. For simplicity, suppose our animal pauses just half the expected time to food. Since the expected time to food right after food is *at least* 2T (the minimum duration of the two fixed intervals), the *first* pause in our chain schedule should be T s. The second pause should be T/2, because the expected time to food after the stimulus change from S_1 to S_2 is T s.

With these values the timing process is both stable and *optimal*, in the sense that our hypothetical pigeon behaves in such a way that food is delivered as soon as the schedule permits: with this pattern he gets food after 2T s every time. But notice what happens if we add one more FI link, making a 3-link chain. If the animal updates his expectation about the time to food very rapidly, from interfood interval to interfood interval, the pattern of pausing is no longer stable at the original values, and not optimal. The sequence of pause values and times-to food is shown in Table 1, which shows hypothetical performance on a 3-link FI 100 schedule.

Table 1
Behavior on a 3-link FI 100 s Chain

		Pause Fraction: 0.5		
Cycle	Stimulus	Expected Time to Food	Pause	Link Length
	S1	300	**150**	150
1	S2	200	100	100
	S3	100	50	100
			Interfood Interval	**350**
	S1	**350**	**175**	175
2	S2	200	100	100
	S3	100	50	100
				375
	S1	375	**188**	188
3	S2	200	100	100
	S3	100	50	100
				388

The key times in Table 1 are in boldface. I assume that in the first cycle the animal expects the minimum possible times-to-food: 300 s just after food in the first link, 200 s at the beginning of the second link, 100 s at the beginning of the last link. But notice what happens in the next cycle: because the pause in the first link (150 s-one half the 300 s expected-time-to-food at the onset of S_1) was longer than the 100 s FI value, the total interfood interval is increased from 300 (the minimum prescribed by the three FI 100 links) to 350 s. In response to this, the first pause in the second cycle increases to 175 s, which means that that cycle is still longer than the preceding one: 375 vs. 350 s. The third cycle is longer still. It is easy to show that the process illustrated in Table 1 stabilizes with an initial pause of 200 s, giving a total interfood interval of 400 s-which is suboptimal because 100 s more than the minimum permitted by the schedule.

The equilibrium interfood interval deviates still more from the minimum permitted by the schedule when we add additional links, because the animal begins to show pauses longer than T (100 s in the example) in links after the first.

We know that proportional timing is a very reliable feature of performance on periodic-food schedules. We have just shown that it explains one characteristic of chain schedules: the excessive pausing in early links. But our explanation still lacks something, because it makes identical predictions for chain and tandem schedules-until the number of links increases to the point that pauses greater than T occur in links after the first (such pausing is possible only on the chain, where the transitions from one link to the next are accompanied by stimulus change). Moreover, the simple proportional timing process always stabilizes at a finite value: timing alone cannot predict the *extinction* that is such a dramatic feature of performance on multilink chain schedules. Animals also do not respond in the rigid way implied by a constant pause fraction: pause duration varies somewhat from link to link. Real animals also do not adapt to changes in interfood interval as rapidly as our ideal pigeon, and their average pause may well be less than 50% of the typical time-to-food. All these deviations from the simple model tend to mitigate its predictions, to shorten postfood pausing on multilink chains. Nevertheless, in practice pigeons, at least, are unable to maintain their responding on chain schedules of more than five or six links. What is missing?

What seems to be missing is a factor that reflects whether or not food actually occurs in the presence of a stimulus: in a tandem schedule, food does in fact occur in the presence of the stimulus; but in a chain, food only occurs in the presence of final stimulus in the chain. The pairing between a stimulus and food contributes a positive factor for performance on the tandem schedule that is lacking for stimuli other than the terminal one on chain schedules. The absence of this factor, plus remoteness from food, produces extinction in the earliest component of a multilink chain. The fact that food actually occurs in the presence of the stimulus accounts for the maintenance of behavior on multilink tandem schedules and on chain schedules in which the components are scrambled from cycle to cycle.

This analysis, and the weak effects of the contingency between response and conditioned reinforcer, de-emphasize the role of conditioned reinforcement as a "strengthener" on a par with food reinforcement. It seems that behavior on chain schedules is determined largely by pairing with food (or its absence) and by temporal proximity to food-as it is on fixed- and variable- interval schedules, and multiple schedules.

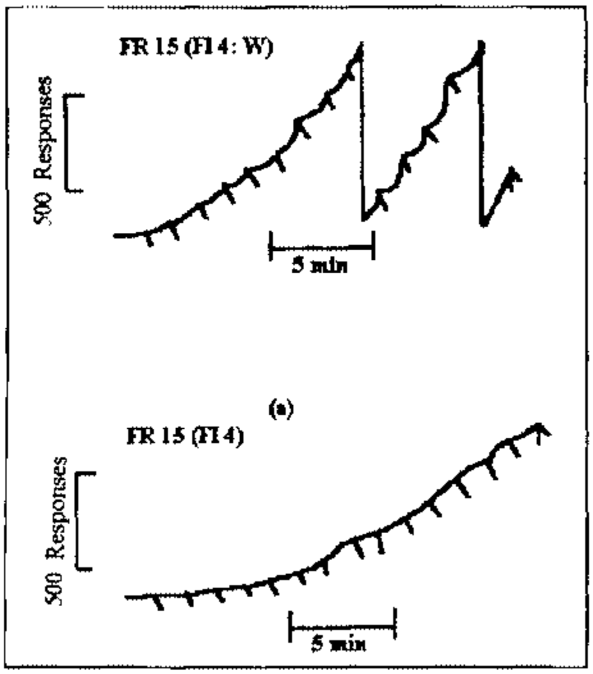
The higher response rate in the final link of a chain schedule, compared with the corresponding tandem, also reflects the allocation of terminal and interim activities. The process is the same one proposed as one explanation for behavioral contrast (Hinson & Staddon, 1978). Interim and terminal activities are segregated into early and late links on chain schedules, because the early stimuli are never paired with food; interim and terminal activities are less-well segregated on tandem schedules, where only (postfood) time distinguishes early and late "links". Hence, terminal responding suffers less reduction from competition, and therefore occurs at a higher rate, in the late links of the chain.

It is also worth noting that a reliable temporal pattern for well-trained animals on multiple schedules, high response rates in anticipation of a "rich" component, and low rates in anticipation of a "lean" component, follows from the kind of timing process described in Table 1. Response rate is high just before the transition to a rich component because expected time to food is short then, low before a transition to a lean component because expected time to food is long (cf. Williams, 1979).

Second-Order Schedules: Stimuli as Impediments to Memory

Stimuli that signal delayed reinforcement facilitate the acquisition of behavior because they aid recall. Figure 5 shows the results of a procedure that is superficially similar to a chain schedule where response-produced stimuli seem to facilitate responding by interfering with recall. The top cumulative record shows the relatively high response rate and scalloped pattern produced in a well-trained pigeon by splitting up a fixed-interval 60-minute schedule into 15 fixed interval schedules with 4-minute components, each terminated by a response-contingent brief (0.5-second) stimulus. Food follows (that is, is paired with) the last such brief stimulus in the 15-component cycle. This arrangement is called a *second-order* schedule. Second-order schedules differ from chain schedules in that the component-change stimulus is brief, and always the same. The bottom record in Figure 5 shows behavior on a comparable tandem schedule. The tandem schedule performance shows that without the brief stimuli response rate is very low, and in the absence of time markers other than food there are no 4-minute scallops within the longer 60-min interfood interval.

Figure 5



Top: Cumulative record from a well-trained pigeon on a second-order FI 60 min (FI 4-min) schedule. Every 4 min a peck produced a brief (0.7 s) stimulus on the response key (indicated by the blips on the record). At the end of the 15th stimulus, food was presented. Bottom: Performance on the same schedule with no brief stimulus presentations (tandem; from Kelleher, 1966).

The brief stimuli in this experiment seem to act by interfering with the animal's recall for the most recent food delivery. There are data from experiments with interpolated stimuli showing that the pattern of pausing following a stimulus that signals a delay to food (on so-called *primed* fixed-interval schedules) depends upon the animal's ability to recall the most recent time marker (cf. Staddon, 1974). The most reliable time marker in the second-order schedule shown in Figure 5, is food delivery. However, food is temporally remote (the interval is 60 min long) and the intervening time is filled with periodic brief stimuli. Moreover, the last brief stimulus is also a potential time marker because food is always preceded by a brief stimulus 4 minutes earlier. The brief stimulus is less reliable than food as a temporal predictor-food only follows one of the 15 brief stimuli-but it is a great deal closer in time: when food occurs after a brief stimulus it does so after only 4 min, as opposed to 60 min after food. The greater validity of food as a time marker seems to be outweighed by the greater frequency and closer proximity-to-food of the less-valid brief stimulus. The brief stimulus seems to retroactively interfere with recall for food delivery as the effective time marker. Because the animal does not know where he is in the interval he must treat each brief stimulus as if it were the last, hence anticipates food at each stimulus presentation. Since a brief stimulus occurs every 4 min, response rate is naturally higher than on the tandem schedule, which is treated much like an FI 60-min. The effects of the second-order stimuli can be so large (when the signaled time-to-food is only a few seconds) that the term *quasi-reinforcement* was once suggested for them (Neuringer & Chung, 1967; Staddon, 1972).

The effects of brief stimuli were at one time attributed to the pairing of the final stimulus with food. Subsequent work has shown that pairing seems to be important only for the initial acquisition of the pattern of responding (Squires, Norborg & Fantino, 1975; Stubbs, 1971).

In summary, conditioned reinforcers seem to aid the acquisition of behavior because they act as aids to memory, that is aids to solving the assignment-of-credit problem, and as signals for food. They aid the maintenance of behavior under some conditions because they signal conditions of more-frequent reinforcement-but they also impair behavior on extended chain schedules because early links in the chain are both remote from, and unpaired with, food. Under other conditions, second-order stimuli can maintain behavior by confusing the animal: by interfering with the animal's ability to recall the time of food delivery.

I earlier described the "reinforcing" effect of the stimulus in hill-climbing terms: the pigeon pecks the White key to produce a Green key, because Green is associated with a higher frequency of food reinforcement than White. Will animals only work for "good news" —a situation clearly better than the current one-or will they work for *information*, good news or bad, about the conditions of reinforcement? This issue has been pursued by several researchers beginning many years ago with experiments on what was called *observing behavior* (Wyckoff, 1952).

Observing Behavior: Information or Conditioned Reinforcement?

In an observing-behavior experiment, pigeons are offered the opportunity to produce a stimulus that tells them whether a reinforcer is likely or not, but has no effect on the actual availability of the reinforcer. Responding sustained by the production of such an *informative stimulus* is termed *observing behavior.* For example, suppose that food for hungry pigeons is scheduled on a variable interval 60-second (VI 60) schedule for pecking on the Left key, which is normally White. Pecks on the Right key have no effect on food delivery, but turn the Left key Green if the VI is due to make food available within the next 30 seconds. Under favorable conditions, pigeons will soon learn to peck the Right key. Does this indicate that "information" is reinforcing in and of itself, or can we deduce this result from familiar principles? The answer is that we do not need to postulate a desire for information on the part of the bird. The delay to food delivery in the presence of Green is less than its delay in the situation as a whole. Hence, pigeons should peck for Green on the Right, even though the observing response has no effect on the overall rate of reinforcement. According to our hill-climbing idea, responding is maintained on the observing key by an overall reduction in the delay to reinforcement. The reduction is only apparent, of course, but the pigeon is not in a position to know that.

A better test for the information idea is a procedure where the Left key is normally White, as before, but a peck on the Right, "observing" key turns it Green if food is *not* to become available in the next 30 sec. This "bad news" procedure gives the bird just as much information as the "good news" procedure just described. But pigeons will not peck the "observing" key under these conditions because it is associated with a delay to reinforcement longer than the overall average. This preference for "good news" is the same as the preference for positive hits in the human experiment described in the previous chapter. It also reflects the underlying hill-climbing process discussed elsewhere: Organisms act in ways that lead to an improvement in their overall situation. Pigeons, at least, have little use for information in the abstract (cf. Fantino, 1977).

Proportional Timing and Ratio Schedules

We have seen how proportional timing seems to be the major factor in performance on fixed-interval, and chain FI, schedules. We will see in a moment how it can help us understand the complex patterns of data on choice procedures. It may also underlie behavior even on procedures that seem to have nothing whatever to do with timing, namely *ratio* schedules. I will pursue the argument with the aid of our "ideal pigeon". I use the term *ideal pigeon* in the way that physicists use the term *ideal gas.* An ideal gas is one that obeys the gas laws perfectly, as real gases do not. Nevertheless, the ideal is close enough to the reality to provide interesting predictions. Our ideal pigeon obeys the principle of proportional timing perfectly: he always waits a fixed fraction of the expected time to food delivery before responding; and when he responds he does so at a steady rate until the next food delivery. I will also assume (following the argument in Table 1) that he updates his estimate of the expected time each time he gets food. We will see in a moment that the ideal pigeon mimics the behavior of real pigeons in very many concurrent chain experiments. What will such a predictable bird do on when he gets food for every peck, i.e., on a fixed-ratio 1 schedule?

The first food delivery doesn't tell him much because he has no time marker (other than the beginning of the session) to tie it to. But the second food delivery occurs time t1 after the first, where t1 is just the time between the first peck and the second. Now he has a time marker food (or a peck, since they occur at the same time) and an estimated time-to-food-t1. Proportional timing says that his next peck will therefore be after a time t2 that will be *shorter* than t1, because the pause is always a fraction of the expected time to food. By the same process, the next pause, t3, will be shorter still, and so on in a positive-feedback process that will soon have the animal responding as rapidly as possible. Rapid responding is of course the main characteristic of performance on ratio schedules.

Suppose now we slowly increase the ratio size, first two pecks to food, then three, and finally ten, say. Now we have two time markers: food, which signals a relatively long time to the next food (because the animal must complete the 10-peck ratio in between); and the peck, which still signals a short time to the next food opportunity (we assume that the animal cannot count, so cannot tell one peck within the ratio "run" from another). The very short minimum peck-food time implies that once the animal begins to respond he should continue to respond at a high rate. But the longer food-food time implies that the *first postfood peck* will be delayed, i.e., the animal should develop a postfood *pause*. The fixed ratio begins to look like a two-link chain schedule: the first link is the time between food and the first peck; the second link is the ratio "run", once pecking has begun. Moreover, the larger the ratio size (the longer the second link), the longer the pause should be. Postfood pause is indeed directly related to ratio size on fixed-ratio schedules.

How long can we continue this process; that is, how large can we make the ratio before behavior collapses and the bird quits responding? Here the simple theory is less accurate, but it is at least possible to calculate how the postfood pause should increase with ratio size. Suppose the animal always waits *half* the expected time to food, i.e., E/2, and suppose that he cannot peck faster than one peck every t sec. If the ratio size is N, the time taken up by each ratio "run" is therefore Nt sec. Hence, the expected time to food, E, is just t_p Nt, where t_p is the postfood pause. But, by our hypothesis, the postfood pause is always half the expected time to food, i.e., $t_p = E/2$. Thus, we have two equations to solve to find the actual value of the pause:

$$E = t_p + Nt, \text{ and}$$
$$t_p = E/2.$$

We just need to eliminate E from these two equations to find t_p as a function of N. The answer is $t_p = Nt$, i.e., the pause is proportional to the ratio value as the verbal argument led us to expect. In general, if the pause fraction is A, where A is between 0 and 1, the relation is $t_p = ANt/(1-A)$.

Experimental results also show that the pause fraction, A, depends upon the size of the food delivery the animal gets: the larger the food delivery, the smaller the pause fraction (Powell, 1969). We will have use for this relationship between pause fraction and reward magnitude later, when we come to discuss self-control. This proportional-timing process is stable no matter how high the ratio, the ideal pigeon should still continue to respond so long as our assumption that the animal cannot discriminate one peck from another holds up. But at very large ratios, the first peck in the ratio "run" occurs a long time after food and may perhaps become discriminable from later pecks. But this first peck signals a long time to food, so that a pause may begin to develop after this peck just like the pause following the onset of the second link in a multilink chain schedule. This pause lengthens the total time between food deliveries, and thus the postfood pause. *Ratio strain* is the name given to the pausing that begins to occur during the ratio run at high ratio values. Ratio strain is usually a sign that the animal is almost at its limit of ratio size and further increases are likely to lead to extinction. It is presumably the cumulative effects of the development of pausing during the ratio run that lead to the eventual breakdown of ratio-schedule performance at high ratios. Other processes, the sheer effort involved in a long run of pecking, for example, may also contribute (I discuss explicitly the hypothesis that effort is a factor in schedule performance later in the chapter).

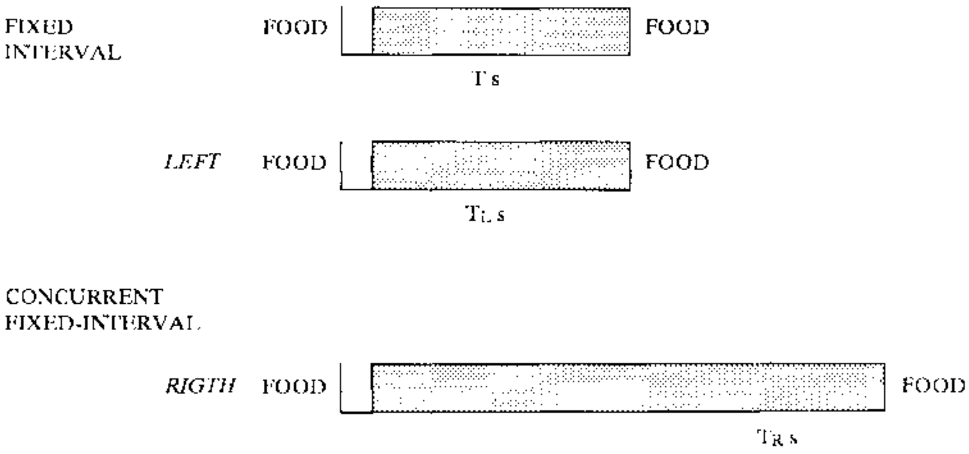
What about *variable*-ratio schedules? Here the positive-feedback process that produces high response rate operates just as it does on fixed-ratio. But food delivery is no longer a reliable signal for a long expected time to food, because some ratio values are very small. Hence, proportional timing implies no postfood pause. Pausing does not in fact occur on VR schedules, although response rate is indeed high, just like fixed-ratio.

Thus, the main properties of responding on both fixed- and variable-ratio schedules, the high "running" rate and the FR postfood pause, are directly traceable to the proportional timing process.

Delayed reinforcement and choice

Chain schedules have been used extensively in choice procedures intended to discover quantitative principles of reinforcement. Some of these procedures are complicated, but the basic design of all of them-and many experimental results-can be understood quite easily by beginning with an idealized two-choice procedure. Look at the FI schedule diagrammed in the top panel of Figure 6, which shows how the ideal pigeon will respond during a single interfood interval. The bird waits for a fixed fraction, say 1/4, of the time to food before beginning to peck, and then (I assume) pecks at a constant rate until food is delivered at the end of the interval. The steady rate of pecking is represented by the filled area that begins T/4 s into the T s interval.

Figure 6



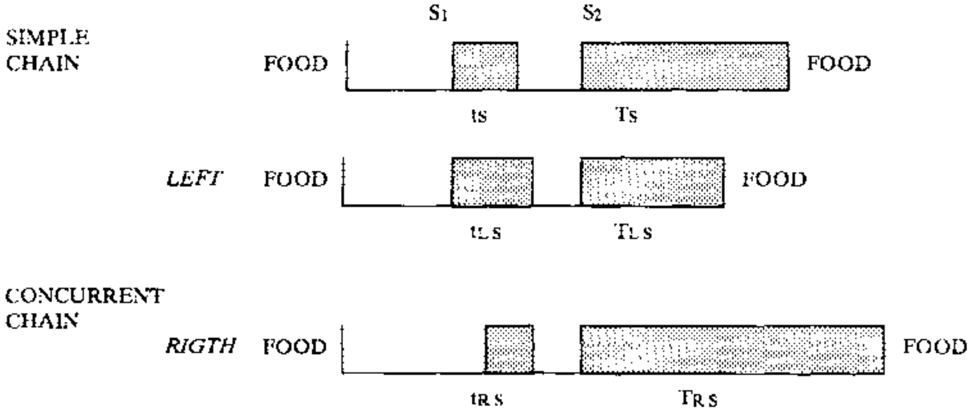
A single cycle of simple (top panel) and concurrent (bottom panel) fixed-interval schedules.

Now look at the bottom panel, which shows a *concurrent* FI FI schedule: two response keys, Left and Right, each of which allows access to its own, independent FI schedule. The two schedules are not of equal length: one is T_L s, the other T_R s and $T_R > T_L$. How will the pigeon respond in this situation? The answer tells us a great deal about how animals treat much more complex procedures.

The simplest assumption we can make is that our ideal pigeon treats each key in the concurrent situation in exactly the same way he would treat it in isolation. That is, he waits $T_L/4$ s after food before beginning to respond on the FI T_L key, and TR/4 s before beginning to respond on the FI TR key (since the pigeon cannot respond at the same time on both keys, there will inevitably be *competition* for available time after $T_R/4$ s, when the bird must peck on both keys; I will ignore this factor for the moment since it affects both keys equally). But there is an ambiguity here: Obviously the animal will actually receive food earlier on the Left, FI T_L, key than on the Right, FI T_R, key. There are now two ways that the procedure might work: (a) the FI T_R timer could continue even after food for a Left response, until a peck eventually produces food on the Right, or (b) *both* scheduling timers, Right as well as Left, could be reset, so that the two FI schedules are timed from a food delivery produced by either key. The important thing to notice is that this procedural difference makes a great deal of difference to how the animal is likely to behave.

If we adopt programming method (a), and reset the Left FI timer only when food is produced by a Left response, and the Right FI timer only when food is produced for a Right response, then we favor some responding to both keys. The reason is that even if the animal starts out by responding almost exclusively on the Left (short-FI) key, after only a few food deliveries a single peck on the Right will produce food because the Right timer continues to run until food is produced by a Right response. A pigeon with some initial tendency to peck both keys will soon detect this feature, hence will continue to make some responses on the Right as well as on the Left. This mixture of Left and Right responding is termed a *partial preference*. Conversely, if we reset both timers after food, according to method (b), with the response pattern shown in Figure 6 the immediate result will be that the animal only receives food for a Left response because the Left timer always sets up first and food delivery resets both timers. Thus, a Right-key peck will never be followed by food and the Right key will become a stimulus that is never paired with food, so Right responding will eventually drop out entirely. Hence, programming method (b) favors *exclusive choice* of the Left key. Pigeons behave as these arguments suggest, showing partial preferences on procedures programmed according to method (a) and tending towards exclusive choice on procedures programmed according to method (b). We will see several examples of both effects later on.

Figure 7



A single cycle of simple (top panel) and concurrent (bottom panel) chain FI FI schedules.

Let's look at a chain-schedule version of the two-choice procedure. The top panel in Figure 7 shows a single interfood interval on a 2-link chain FI schedule: an initial link of t-s duration, in the presence of stimulus S_1, followed by a response-produced transition to a second link of T-s duration in the presence of stimulus S_2. The two response records show the hypothetical performance of our ideal pigeon, who follows the now-familiar pattern described in Table 1. After the initial time marker (food) he pauses a fixed fraction (say 1/4) of the total time to food (i.e., $(t+T)/4$ s) and thereafter responds at a steady rate until the response-produced transition from S_1 to S_2. When S_2 comes on, he pauses the same fraction of the remaining time to food (i.e., T/4 s). (I ignore the additive effect of the pairing of S_2 with food because that has no bearing on behavior in S_1, which is what we are really interested in).

The bottom panel in Figure 7 shows a *concurrent chain* schedule, which is just two keys, programmed in the same way as the single key in the top panel but with different values for the duration of the second link, which is T_L on the Left and T_R on the Right. The procedure has an additional feature that was not necessary on the single-link concurrent FI FI schedule: when the pigeon successfully produces a transition to *either* second link (i.e., either S_{L2} or S_{R2}), the other key is turned off until after food delivery, when both keys return to their initial-link stimuli. Thus, once the animal has produced a second link, he is committed to it until food is delivered. The programming rule for the t-s timer in the first links is rule (b), above: that is, when food is delivered in either second link, both first-link FI timers are reset.

The response records for our ideal pigeon in Figure 7 show how he will behave on this new procedure. The main result is obvious: he waits much longer before beginning to respond in the first link on the Right, which leads to the longer second link. Because the proportion of time taken up by responding (i.e., the shaded area) is larger on the Left than on the Right, the ideal pigeon will seem to be responding at a higher *rate* in the first link on the Left (which has a short second link) than on the Right (which has a longer second link). Notice that because of the "commitment" feature of the procedure (once a second link is entered the other key becomes ineffective) the animal will continue to get about 50% of his food for pecking on the Right key. Nevertheless, because of the late start on the Right, the proportion of responses is obviously much less on the side leading to the longer second link.

This result some responding on both sides, but more on the side leading to the shorter second link is not surprising, although responding on both sides is not always optimal behavior under these conditions (I return to the question of optimal policy in a moment). Nevertheless, it is in fact what pigeons usually do, although the schedules used are little more complex than the ones shown in Figure 7. In the usual concurrent chain experiment, the delays in the second links are often fixed, as in our example, but the first-link schedules are usually *variable-* rather than *fixed-* interval. The reasons for this are largely historical this is how the first experiments were done (e.g., Herrnstein, 1964) but this difference does not affect our conclusion, which depends only on the animal setting his waiting time in the two initial links to a fixed fraction of the expected time to food.

Optimal Policy on Concurrent Chain Schedules

How *should* pigeons respond on concurrent chain schedules with equal first links? What should they do so as to minimize the average time between food deliveries? The pigeon really only has two options in this situation:

either ignore the key leading to the longer second link, or sample both in such a way as to enter each second link as soon as it is available. I'll call the first strategy *fixate* (on the short-link side) and the second *sample*. It is pretty obvious from Figure 7 that with fixed-duration, equal, first links, that are reset with each food delivery, there is absolutely no reason ever to respond on the side that has the longer second link. But what if the first links are variable rather than fixed-interval? In this case, everything depends on the *relative* durations of the first and second links. The argument is straightforward: For the strategy *fixate* the average interfood interval will be just $t+T_L$, where t is now the value of both first-link VI schedules. But for the strategy *sample* the average interfood interval will be the average of two numbers: a number representing the average time the animal spends in the first link, plus a second number which is the average of the two second-link delays. The first number, when the first links are equal VI's, will in fact be *half* the VI value, since the two VI's are independent and each sets up once every t s. Thus, the average interfood interval under the *sample* strategy is just $(t+T_L+T_R)/2$. *Fixate* is better than *sample* if the corresponding interfood interval, $t+T_L$, is less than $(t+T_L+T_R)/2$, which reduces to the condition:

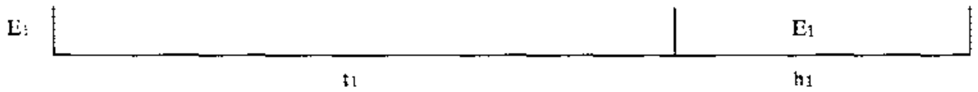
$$t+T_L<T_R$$

In short, the animal should be more likely to fixate as the duration of the first-link VI, t, decreases. This makes perfect intuitive sense: by waiting long enough, the animal can ensure that a single peck on the Right gets him immediately into the second link. The point is that this peck will only be worth making if the delay he must then suffer, T_R, is shorter than the expected time to food for just continuing to respond in the first link on the Left, $t+T_L$.

The interesting thing about this prediction is that this shift, away from fixation towards indifference as t, the first-link VI, is increased, is precisely the pattern that pigeons show on these procedures (Fantino, 1969). I show in a moment that this pattern, and a number of other results from experiments with concurrent chain schedules, is also what we would expect from our ideal pigeon who behaves close-to-optimally under these far from-natural conditions. But of course the ideal pigeon is not optimizing at all: he is simply following blindly the proportional-timing rule, which happens to work pretty well in the situations I have discussed so far. Naturally there are situations where proportional timing does not work so well; and in these situations the ideal pigeon behaves suboptimally. It is interesting that these are situations where real pigeons also fail to maximize their rate of food delivery. I return to these topics in a moment.

Chain Schedules and Natural Foraging. The choice between two chain VI FI schedules is formally the same as the problem of choosing between two different prey types that have different profitabilities. This is a classic problem in behavioral ecology, the problem of diet selection (see Stephens & Krebs, 1986, for an excellent survey of theoretical and experimental work). For example, imagine a predator (a squirrel, say) who has two types of nuts available to him. Let's for the moment assume that both nuts have the same nutritive value (food amount), but that nut A has a thicker shell than nut B, so takes longer to crack (i.e., has a longer *handling time*). To further simplify the situation, I will assume that the two types of nut are equally abundant. When should our squirrel eat either type of nut whenever he encounters it, and when should he just specialize on the more profitable type? (*Profitability* is just the ratio of energetic value, E, to *handling time*, h: profitability = E/h). If each type of nut is encountered on the average every t s, his rate of energy intake if he specializes on B is $R_B = E/(t+h_B)$, where h_B is the handing time for B, i.e., R_B is equal to the energy-per-nut, E, divided by the average time between nuts, which is encounter interval t, plus handling time h_B (see Figure 8). If he generalizes, accepting both A and B, on the other hand, his rate of energy intake will be E, divided by the average encounter time, which is now t/2 (since we now count encounters with either equally abundant type) plus the average handling time, $(h_A+h_B)/2$, so that $R_A = 2E/(t+h_A+h_B)$. The animal should specialize if $R_B > R_A$, which reduces to $t+h_B < h_A$ —which is the same as the condition for the policy fixate in our chain-schedule analysis. The main ecological implication of this analysis is that as food density decreases (i.e., encounter time, t/2, increases) the animal should be less and less likely to specialize and more and more likely to accept any food he encounters, even the least profitable. Numerous experiments have confirmed this qualitative prediction in several species (See Stephens & Krebs, 1986).

Figure 8



A typical cycle of foraging for a food type with profitability E_1/h_1, encountered on the average every t1 s.

A more general diet-selection problem provided the initial impetus for the whole field of *optimal foraging theory*: this is the so-called *optimal-diet* problem (Macarthur & Pianka, 1966). Consider two prey types, one of which is highly profitable (low handling time and/or high energy content) the other less profitable: $E_1/h_1 > E_2/h_2$. These two types have different abundances: type 1 is encountered on the average every t_1 s, type 2 every t_2 s. The question is: Under what conditions should the animal take both types, whenever he encounters them *(generalize)* and when should he just take the more profitable one (*specialize*)? The analysis is similar to what we have just done, but it also allows us to introduce a new technique, the method of *marginal value*. I first derive an expression for the expected food rate under strategy *specialize*. The situation here is just like a simple chain schedule, as shown in Figure 8. It is obvious from the diagram that the expected rate of energy acquisition under *specialize* is just $R1= E_1/(t_1+h_1)$, that is, the energy in one prey item, E_1, divided by the time between items, t1, plus the time it takes to eat an item, h_1.

Generalize is a bit more difficult, because we must somehow average the payoff from both types of items. At first blush, it might seem that our predator should *always* take even the less-profitable item, but that is because we forget what economists call the *opportunity cost* of eating a low-profitability prey item. The opportunity cost is incurred by the item's handling time, h_2: if h_2 is long, then it might be better for our optimal predator to spend the time looking for the more profitable prey, rather than waste time struggling to eat the less-profitable one. The easiest way to see the optimal strategy here is to use this argument, which is known as a *marginal value* argument. The idea is that under most conditions an animal can maximize its rate of return by always picking moment-by-moment the option with the highest expected reward rate. The base comparison here is of course with the *specialize* strategy: if the animal ignores the less-profitable item, he can always get food at a rate $R1= E_1/(t_1+h_1)$. So when he *encounters a less-profitable item*, he should compare R_1 with the expected food rate once he has already encountered the less-profitable item, namely E_2/h_2, the profitability of the lesser food type. This argument yields the counterintuitive result that in a situation like this, the abundance of the less-profitable item (determined by t_2) should have *no effect* on the animal's willingness to take it-only the abundance of the more-profitable item (determined by t1) should have any effect.

Under many conditions this prediction seems to hold up. For example, Krebs et al. (1977), did a laboratory prey-selection experiment with great tits (*Parus major*), small insectivorous European birds. The birds were presented with artificial prey (mealworms) that passed in front of them on a little conveyor belt. There were two types of prey, "bigs" and "smalls"; the experimenters varied the frequency of the "bigs" to see if there was indeed a critical frequency above which the birds would suddenly cease to take any "smalls". They did find a transition, but it was smooth, rather than sudden (you can probably think of many reasons for the gradualness of the change: how well could the birds estimate frequency, for example?). In another experiment, Werner and Hall (1974) looked at bluegill sunfish feeding on *Daphnia* (water fleas) of three different sizes in a tank. There were equal numbers of each size. The more fleas there were in the tank, the less likely the fish were to take the smaller sizes.

Thus, optimal foraging analysis can be extended to make predictions about what animals should do when the two food types have different abundances, or are of different nutritive values. The case where encounter rates are the same, but nutritive values and handling times are different, resembles the *self-control* experiments I discuss later on (see also Box 1). These optimality arguments don't say anything about the mechanisms or rules that underlie the animal's behavior, of course. We will see in the next section that proportional timing, which is the main mechanism that seems to drive choice on these procedures, does not always produce optimal behavior. We can therefore expect optimal foraging predictions to fail under some experimental conditions.

Parametric Effects on Concurrent Chain Schedules

Pigeons will peck a key to produce a stimulus if the stimulus signals a higher rate of food delivery than the prevailing rate. This finding led researchers to assume that the value of a conditioned reinforcer is directly related to the rate of primary reinforcement in its presence. They believed that (a) pecking to produce conditioned reinforcers is maintained by the value of the conditioned reinforcers; and (b) that the value of a conditioned reinforcer is related to the rate of primary reinforcement that it signals.

As you have seen, it now seems unlikely that the *reinforcing* property of conditioned reinforcers is as important to the maintenance of pecking as their proximity to food. Nevertheless, you need to know the theoretical presumptions that have led to a long series of experiments aimed at measuring the value of conditioned reinforcers. The reasoning was along the following lines: Perhaps stimulus A, in whose presence food occurs after T s, is about half as valuable as stimulus B, in whose presence the same food occurs after a delay of only T/2 s. If this is true, perhaps pigeons will work twice as hard to B as for A. How can we verify this prediction, researchers wondered? We might ask pigeons, on a trial-by-trial basis, to choose between two keys, one displaying stimulus A, the other stimulus B. But even the stupidest pigeon is likely to choose stimulus B exclusively. This tells us something about the relative values of A and B, namely that B is greater than A; but it does not tell us *how much* greater. Something more subtle was needed.

Richard Herrnstein (1964) proposed an ingenious solution to the problem of measuring the value of a conditioned reinforcer, based on an earlier result with primary (food) reinforcement. He and his colleagues had already shown that if pigeons are allowed to respond concurrently on two independent variable-interval schedules they will approximately *match* their ratio of key pecks, Right/Left, to the ratio of obtained food reinforcements (if the VI schedules are different), or to the ratio of food magnitudes (if the VI schedules are both the same); see Herrnstein (1970) for a review. If we denote rates of responding on Right and Left by x and y, and rates of reinforcement obtained by R(x) and R(y), the simple *matching law* is just

$$x/y = R(x)/R(y).$$

I will say more about the matching law later in the chapter.

The matching-law result with schedules of primary reinforcement suggested to Herrnstein a procedure to measure the value of conditioned reinforcers: Why not ask pigeons to choose between two *conditioned* reinforcers, each delivered according to the same VI schedule? This is just the concurrent chain procedure we have been discussing, but with VI, rather than FI, schedules in the first link. The matching result implies (Herrnstein argued) that the ratio of key pecks in the first links should provide an accurate measure of the ratio of conditioned reinforcing values of the second links.

In his first experiment Herrnstein found, as he had expected, that the ratio of pecks on the Right and Left in the equal-VI first links matched the inverse ratio of second-link delays: thus, if the second link delays were $T_L = T$ and $T_R = T/2$, pecks Right/pecks Left in the first links were in the ratio 1 to 2. Evidently the value of a conditioned reinforcer is indeed inversely related to the food delay in its presence.

However, subsequent experiments soon showed that this conclusion is valid only under certain very specialized conditions. From a present-day perspective we can see that Herrnstein's result was unlikely to be universal because it rested on a functional argument explaining behavior in terms of its outcome. Like any functional principle such as the principle of reinforcement itself Herrnstein's simple rule took no account of the mechanisms that underlie performance. Just as reinforcers sometimes fail to reinforce (cf. instinctive drift and superstitious behavior), so conditioned reinforcers will sometimes fail to act in the expected way. Just as the failures of the reinforcement principle have told us something about the mechanisms that underlie reinforcement, so the failures of conditioned reinforcement tell us something about how *it* works.

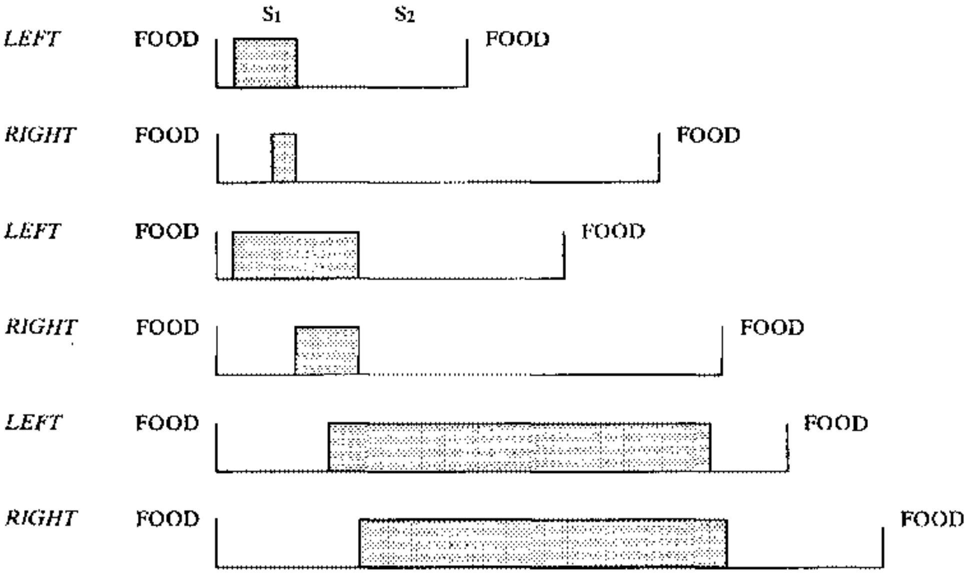
The first experiment to cause problems showed that the quantitative preference for the shorter-second-link key diminishes as the length of the first-link VI's is increased: the longer the first links, the more the animal tends to be indifferent between the two keys (Fantino, 1969). As we saw earlier, this makes intuitive sense: if the first links are very long relative to the both second links, then the animal's optimal policy is to get out of the first link, either link, as soon as possible. It follows directly from the optimality condition for fixation on the short-link side we derived earlier, $t + T_L < T_R$: the longer the value of t, the first-link delay, the less likely the animal is to respond on just one key. This kind of result poses severe, if not insuperable, difficulties for the simple conditioned-reinforcement idea. The difficulties come from the fact that the value of a conditioned reinforcer depends on more than the rate of primary reinforcement that it signals: evidently conditioned-reinforcing value is a relative rather than an absolute property. The difficulties are not insuperable because we can simply revise our functional principle to accommodate this relativity, and numerous attempts have been made to do so. But the cost is greatly increased theoretical complexity, and loss of predictive power.

Over the years several other limitations on Herrnstein's original conclusion have appeared. First-link preference is affected by the absolute (as well as relative) durations of the second links and by the variability of second-link delays, as well as by the absolute value of first-link delays. First-link preference is surprisingly little affected by the *number*, as opposed to the delay, of second-link reinforcers.

I now show how all these effects can be derived from the proportional timing mechanism. The arguments are as follows:

Effect of First-Link Delay. Figure 9 shows how our ideal pigeon should respond on three concurrent chain schedules with the same pair of unequal second link delays, and increasing, but equal, first-link delays. As before, we assume that first-link responding begins when 1/4 the expected-time-to-food has elapsed. The period of responding during each first link is indicated by the cross-hatched region. It is easy to see that the *ratio* of first-link pecks approaches indifference (1:1) as the first-link delays increase.

Figure 9

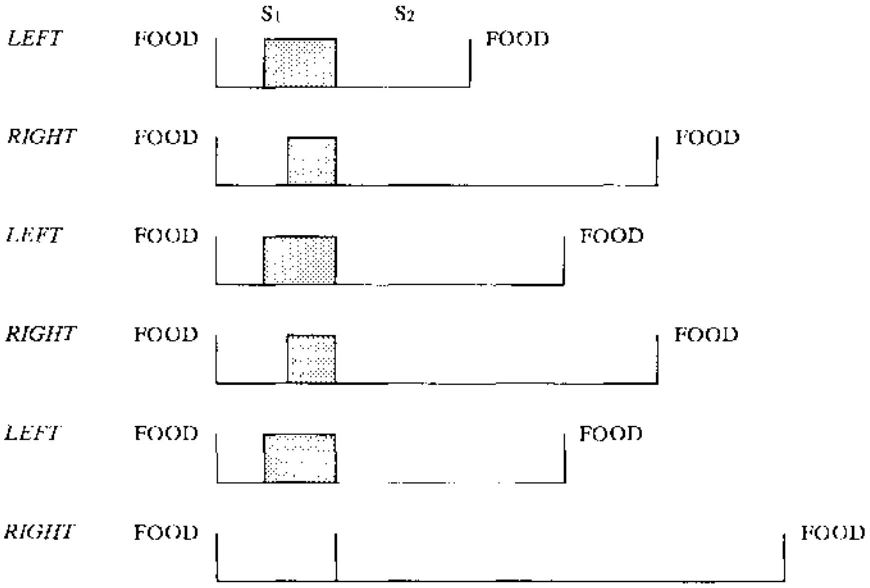


The effect of first-link duration on concurrent-chain-schedule performance. First-link responding begins 1/4 time to food delivery in every case.

Effect of Second-Link Delay. Figure 10 shows how our ideal pigeon adapts to increasing the absolute duration of the two second links, while keeping the equal first links constant. It is easy to see that as the absolute durations of the two second links increase, preference should shift towards exclusive choice. This is basically the same as the previous result: in the first case, the second links were held constant and the first links increased; in this case the first links are held constant and the second links are increased. Real pigeons show both the predicted effects (MacEwen, 1972; see review in Fantino et al., 1972).

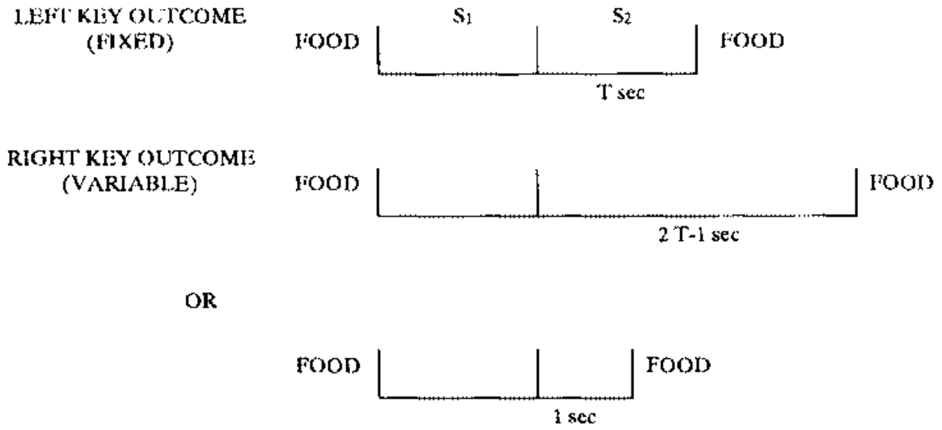
Effect of Variable vs. Fixed Second-Link Delay. Figure 11 shows the time relations in a concurrent-chain experiment with equal first links. The second link on the Left is a constant delay, T. On the Right *two* different delays, 1 or 2T-1, averaging T, occur with equal probability. Numerous experiments have now shown that pigeons have a strong preference for the variable alternative, i.e., the one that leads to either a short or long second link with equal probability. Our ideal pigeon also shows this effect because he begins responding in the initial link in anticipation of the first food *opportunity* (providing that opportunity is not too infrequent). Since the variable side provides the first opportunity, responding should begin much earlier on that side. The ideal pigeon becomes indifferent between the two alternatives only when the fixed alternative is almost equal in value to the shorter of the two variable alternatives (rather than equal to their mean, as the optimal policy requires).

Figure 10



The effect of second-link duration on concurrent-chain-schedule performance. The ratio of second-link delays is constant in each case. First-link responding begins 1/4 time to food delivery in every case.

Figure 11



Concurrent-chain procedure with variable vs. fixed second links. The two *variable RIGHT* second links occur with equal probability and have the same average duration as the *fixed LEFT* second link.

James Mazur (1986) developed a negative-feedback or *titration* procedure that allowed him to estimate the point of indifference in procedures like this. His schedule is complicated, but its essential feature is that if the animal shows a preference for the *variable* side, the duration of the *fixed* second-link delay is reduced; conversely, if he prefers the *fixed* side, its duration is increased. In this way the animal is driven towards indifference. Mazur found that given a choice between a *variable* alternative with 0 and 20 s delays, pigeons will become indifferent only when the *fixed* delay is set to a little over 1 s, a strikingly large departure from the optimal policy, which implies indifference when the fixed delay is 10 s. In another confirmation of the proportional-timing hypothesis, Gibbon, Church, Fairhurst and Kacelnik (1988), who have also developed an elegant quantitative theory of scalar timing, have shown that pigeons greatly prefer a *distribution* of variable-length delays where short delays predominate to a distribution with equal mean and variability where long delays predominate. The Gibbon et al. result shows quite clearly that the pigeons' behavior in Mazur's experiment does not in any way imply a preference for *variability* as such. The birds prefer the short-delay distribution because the expected time to the first food opportunity is shorter than in the long-delay distribution (cf. Williams', 1979, data on following-component effects in multiple schedules, discussed earlier).

Effect of Number of Second-Link Reinforcers. Figure 12 shows a procedure that one might expect to produce a strong Right preference. First- and second-link delays are the same, but on the Right two reinforcers are delivered in close succession before the first link reappears. Optimal responding here implies a strong Right preference, but our ideal pigeon shows an outcome close to indifference, because the *first* food opportunity is at the same delay on both sides. Real pigeons are also almost indifferent between the two sides on this procedure. The second second-link food delivery does have a slight effect, smaller the larger the delay between it and the first (Mazur, 1986; Moore, 1982).

Figure 12

Concurrent-chain procedure in which the second link on one side gives two food deliveries for every one on the other side.

Conclusion

The mechanisms that underlie performance in concurrent chain schedule experiments, turn out to be two in number, both quite simple: (a) an effect due to the pairing of a primary reinforcer (food) with a stimulus, and (b) an effect due to the delay between a time marker and the primary reinforcer. Pairing adds value and reinforces to the extent that the rate of food delivery in the presence of a to-be-produced stimulus is greater than food rate in the current stimulus (hill-climbing). Reinforcer delay acts through a timing process that causes the animal to wait an approximately fixed proportion of the expected time to food before responding (proportional timing). The effects on well-trained pigeons of many experimental manipulations of initial- and second-link length, of second-link variability, and of number of second-link reinforcers are all explicable by the effects of delay alone.

This simple summary nevertheless omits a number of secondary effects. This is an active research area and the issues, and experimental procedures, can become complex. There is some evidence for effects that cannot be reduced to our two processes: Pigeons seem to anticipate more than just the next food delivery, for example; and very infrequent food opportunities are less effective than 100% food. But it is fair to say that these effects are less powerful and reliable than the ones I have described. Pairing, and proportional timing, seem to be by far the most important processes involved in the performance of pigeons on chain schedules.

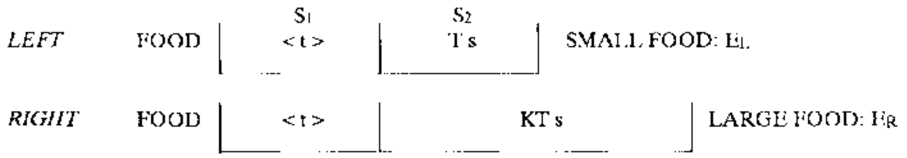
It is noteworthy that these mechanisms, simple though they are in outline, are sufficient to produce behavior that is close to optimal under very many conditions. It is also noteworthy that the situations in which behavior is optimal have close parallels in nature (simple concurrent chains resemble the optimal diet problem, for example), whereas those in which behavior is suboptimal have no obvious natural parallel (concurrent chains with fixed vs. variable second links or links with different numbers of reinforcers, for example).

The Self-Control Problem

People's inattention to delayed consequences poses a chronic problem for public policy makers. People smoke, drink and take drugs for the present pleasure they give, and ignore the future pain, which is not only delayed but is often only probabilistic: Cigarette smoking gives *some* people cancer; it may not give *me* cancer. Individual economic decisions also are often made unwisely because of a failure to evaluate future events realistically: we may impulse-buy now, and fail to pay our rent or insurance later. These obvious practical applications have maintained interest in self-control as a separate problem area, even though, in the present context, it is but one of many possible variations on the concurrent chain procedure.

A typical self-control experiment is depicted in Figure 13 (Green & Snyderman, 1980). It is a minor variation on the second-link-duration problem depicted in Figure 10. The only difference is that the procedure compares second links that differ not just in the delay to food, but also in the *amount* of food. A successful peck on the Left, say, produces a small reward, E_L, delivered after a short delay, T; a successful peck on the Right produces a larger reward, E_R, delivered after a longer delay, KT (i.e., $K < 1$; let $E_R/E_L = A > 1$). The presumption is that under most conditions it pays to show self-control and choose the longer delay to get the larger reward rather than being "impulsive" and choosing the short-delay alternative. We will see in a moment that the optimal policy depends on a number of things: sometimes impulsiveness pays.

Figure 13

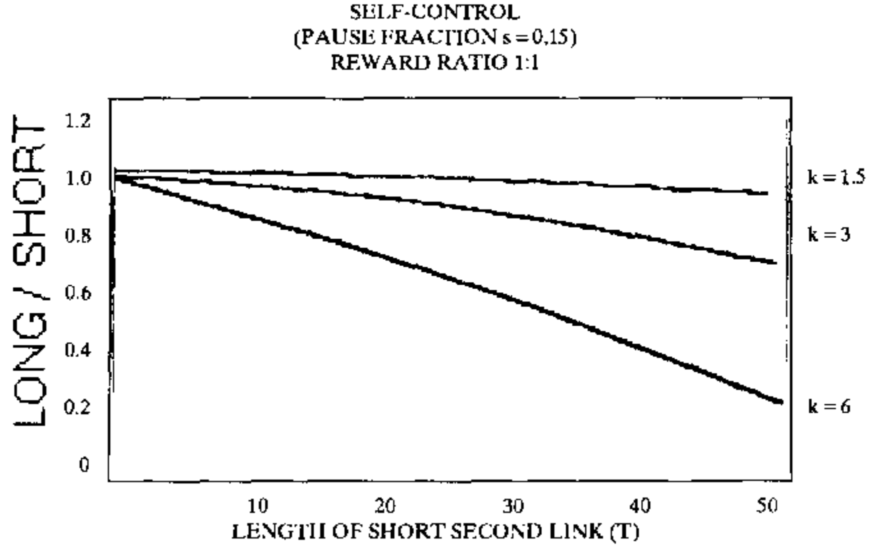


The self-control problem. S_1 is present during the initial equal delay (usually equal VI t s schedules). A successful response produces one of the two delays, and associated food amounts in a single second link.

Proportional timing seems at first to aid us little in predicting the results of these experiments, because the small reward always comes sooner than the larger. Strict application of the rule would have the ideal pigeon always showing impulsiveness, i.e., picking the short-delay alternative. But of course this is what he is likely to show so long as the two reinforcers are of equal size and the initial link is not too long. To account for different results, therefore, we must include a factor that biases the animal in favor of the large-reinforcer choice. The simplest possibility is just that the amount of food affects the proportion of time the animal waits before beginning to respond (cf. Powell, 1969, first encountered in the earlier discussion of fixed-ratio schedules). If he waits 2 s before beginning to respond in anticipation of a small amount of food delivered after 10 s, he might wait only 1 s for a large amount delivered with the same delay: increasing the amount reduces the pause fraction. We will see that this assumption seems to account for most of the results from the extensive manipulations of second-link delay in the Green and Snyderman (1980) experiment.

Green and Snyderman looked at three ratios of long to short second-link delay (I denote the short link by T, the long by KT). The three values for K were 1.5, 3 and 6. For each value of *relative* delay, K, they looked at several values of T, the *absolute* delay. They knew the earlier results on varying second-link delay (T) with *equal* rewards on both sides; they were interested in seeing how these results are changed by giving *unequal* reward amounts for the short and long delays. We have already seen that when the rewards are equal, increasing T always produces a shift of preference towards the shorter link. Our ideal pigeon shows exactly this pattern of results in Figure 14, using a pause fraction of 0.15. The downward slope of the curves indicates that as absolute second-link duration, T, increases, preference shifts away from the longer-link alternative at all three K (relative delay) values. (The formulas i used are described in Box 1, which can be skipped by the nonmathematical).

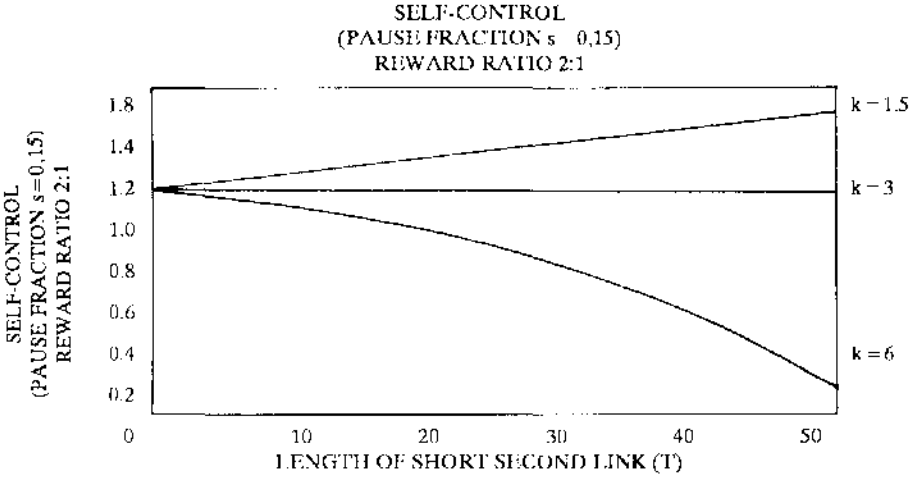
Figure 14



Predictions of proportional timing for the concurrent chain procedure with constant- ratio second links. Graph shows ratio of first link responses (LONG/SHORT) as the absolute value of the second links (T) is varied, at three long-short ratios (K).

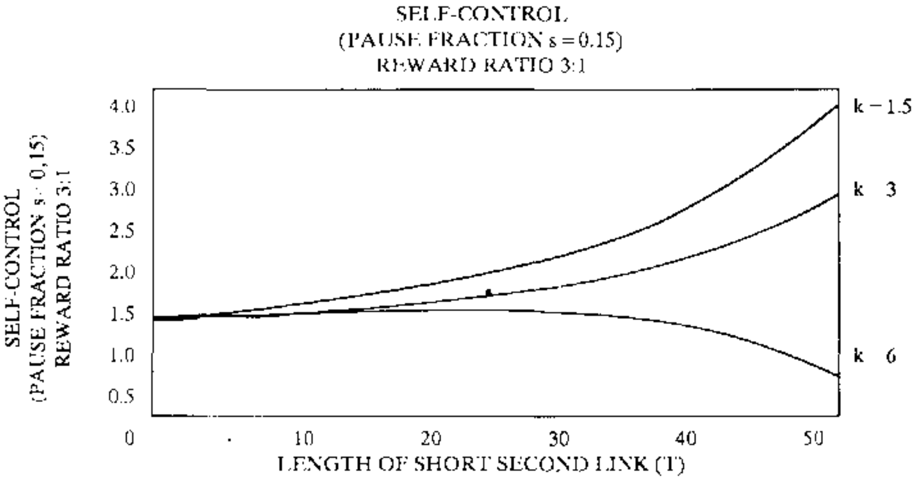
Green and Snyderman found that when the shorter link led to 2 s access to food and the longer link to 6 s access, the results were slightly different from those shown in Figure 14: at the smallest K-value, 1.5, preference for the longer link *increased* as T was increased; at the other two K-values, however, preference shifted towards the shorter link, as it does when both rewards are equal. The ideal pigeon also shows this pattern. in Figure 15, assuming that the pause fraction in anticipation of the larger reward is half the pause for the smaller. The prediction on the assumption that the large-reward pause was one third the smaller is shown in Figure 16: now preference shifts toward the longer side at two of the three K values. There is some reason to believe that the effectiveness of food rewards is less-than-proportional to their physical magnitudes (see later discussion of utility functions), as we had to assume to obtain correct predictions of Green and Snyderman's results (The same assumption is also necessary to reconcile these data with the optimal policy: see Staddon [1980], p. 239).

Figure 15



Predictions of proportional timing for the "self-control"concurrent chain procedure with constant-ratio second links and 2:1 long:short reward ratio. Graph shows ratio of first link responses as the absolute value of the second links (T) is varied, at three long-short ratios (K).

Figure 16



Predictions of proportional timing for the concurrent chain self-control procedure with constant-ratio second links and 3:1 long:short reward ratio. Graph shows ratio of first link responses as the absolute value of the second links (T) is varied, at three long-short ratios (K).

Conclusion. Proportional timing, in the form of our ideal pigeon, seems to handle the rather complex pattern of experimental results on self-control procedures, as it does for other concurrent-chain schedule data. There are of course many differences between real and ideal pigeons: real pigeons show considerable variability in timing, although average pause is related to minimum time-to-food in the same way as for our ideal bird; real pigeons don't adjust their waiting time just interval-by-interval, but also day-by-day-they show long-term as well as short-term effects; and limitations on memory mean that some time markers are better than others. These differences mean that some of the most direct predictions from our ideal pigeon that pausing, not just average response rate, should vary in the first link of concurrent chain schedules, for example, are not as well supported as the more indirect predictions about patterns of results (although, to be fair, no one seems to have looked carefully at anything *but* average rate).

There are also other, more elaborate descriptive theories of performance on chain schedules (see, for example, Fantino's *delay-reduction theory*, e.g., Fantino & Abarca, 1985; the scalar timing of Gibbon at al., 1988; Killeen's *incentive* theory, e.g., 1985). The theory of Gibbon et al. is equivalent to proportional timing but makes additional assumptions about the stochastic properties of the timing process; delay-reduction under many conditions reduces to the optimal policy; incentive theory is also a delay-based theory that makes many of the same predictions as our simplified ideal pigeon. Many details remain to be resolved; but the overwhelming importance of reward delay, in self-control and other chain-schedule procedures, is now very well established.

Although interest in the self-control problem is fueled by its obvious similarity to human situations in which delayed consequences sometimes fail to affect people's behavior as we feel they should, we cannot generalize directly from the animal results to the human situation. The reason goes beyond these particular experiments. It is just that what generalizes (if anything) is not the *particular result* of any animal experiment, but the *principles* and *mechanisms* that underlie the result. It is not sufficient to show that pigeons or rats show self control or impulsiveness under this or that condition. We need to know *why* they show these effects before we can confidently apply what we have learned to the human case.

Although the theoretical problems in this area are far from settled, our ability to predict the major patterns in the Green and Snyderman experiment, one of the most careful and elaborate done to date on this problem, from the idea of proportional timing supports the hypothesis that this mechanism underlies these effects. But proportional timing cannot operate until the animal has had opportunity to learn about the time intervals involved which means repeated cycles of exposure to each alternative and its outcome. But this is almost never possible in those cases of most urgent interest to public policy such as the smoking habit or prudent personal spending. We cannot run people through several lifetimes so that they can compare at first hand the consequences of smoking vs. abstinence, or profligacy vs. prudence. We may therefore have to look elsewhere than laboratory self-control experiments with people or pigeons for public policy recommendations on how to eliminate bad habits with long-delayed consequences. The major problem may not be people's assessment of delayed consequences as much as their assessment of things about which they have been told, but which they have never actually experienced.

What we *can* take away is the overwhelming importance to organisms of strongly valued events, such as food for a hungry animal. These choice situations may not, from the pigeon's point of view, involve any real choice at all. The pigeon is not much distracted by our colored lights and response contingencies remote from food. He keeps his eye on the ball and for the most part allocates his behavior in relation to the ultimate goal, food, rather than the intermediate ones we have set up for him. People, on the other hand, all too often mistake the pointing finger for the moon and toil at subgoals while losing sight of the ultimate objective. Whether this is a reflection of the greater cognitive complexity of human beings, which allows them better to compartmentalize their tasks, or whether it is simply because their goals (at least in psychology experiments) are not as important to them as food for an underweight pigeon, is not clear. Because we cannot ethically manipulate human motivation to the extent that we can alter the hunger of a pigeon, we cannot really answer this question directly, with human subjects. We therefore have little choice but to continue to pursue through animal experiments the problem of how organisms choose between highly attractive alternatives.

Implicit choice: the allocation of behavior

People and animals choose things that have value, and they prefer things of higher value to things of lower value. Thus, the study of choice is also the study of reinforcement. Yet organisms do things all the time, and many of the things they do seem not to be directed at any particular goal; much behavior seems to exist "for its own sake" and not for its consequences. Perhaps there are conditions under which *all* activities (and not just *consummatory* activities like eating or copulation) are reinforcing? This is what David Premack proposed some years ago. He argued that essentially all activities have some value, and that the necessary and sufficient condition for positive reinforcement is just the opportunity to exchange a less-valuable activity for a more-valuable one. His work has led to major advances in our understanding of the functional basis for reinforcement. Extensions of it, in the form of economic and optimality accounts, have revealed unsuspected relationships between implicit and explicit choice, i.e., between how animals allocate their time, and how their choice behavior is guided by reinforcement. I begin with a discussion of implicit choice and the allocation of behavior.

The Premack Principle

Imagine an animal such as a rat in a semi-natural environment in which it can do several things each of which we can record and make available or unavailable. There are two main kinds of experiment we can do with such a situation: (a) Add or remove the opportunity to do various things and look at the effect on the amount and temporal pattern of the remaining activities. Or, (b) impose *contingencies* between pairs of activities and see if the operant response increases in rate. For example, make the animal run in order to get access to water, and see if he runs more. David Premack was the first to see a connection between the pattern of activities under free conditions measured in experiments of the first type and the effect of making one activity contingent on another the second type of experiment. In studies with rats and monkeys, he showed that the effect of making access to one activity contingent on performance of another depends upon the levels of the activities under unconstrained conditions. Let's examine one of Premack's experiments and then describe the ways in which the imposition of a reinforcement contingency can be expected to change the levels of the activities involved.

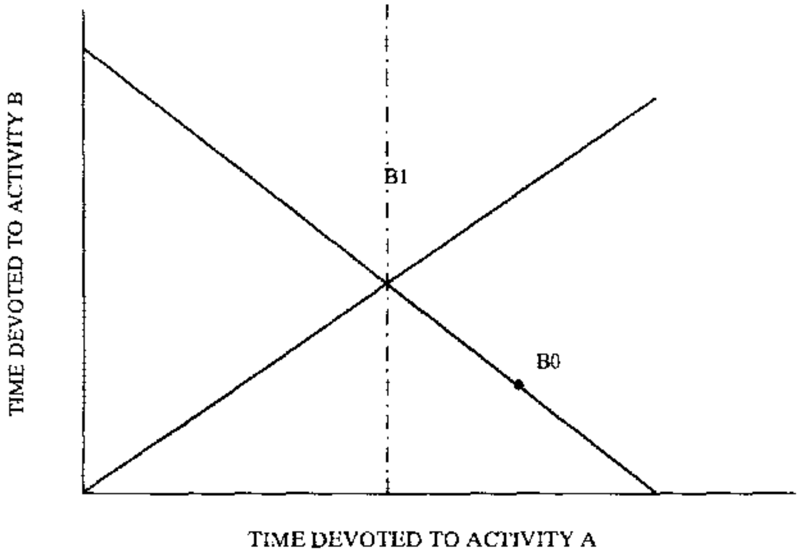
In this experiment, Premack (1965) studied the reinforcing relations among the activities of Cebus monkeys. The monkeys were in individual cages and there were four things that they could play with: a lever (L), a plunger (P), a hinged flap (F), and a horizontally operated lever (H). Access to each of these things was controlled by the experimenter. Premack's idea was that "reinforcement" is not an absolute property of a consummatory activity, such as "eating food", but is just the relation between a more probable activity and a less probable one. In other words, it is the fact that eating, for a hungry animal, is more probable than (say) lever pressing that makes eating reinforcing, not something special about eating itself. Premack's experimental situation was more effective for testing this idea than the usual Skinner box, both because more activities were available, and because different animals showed different preferences for the different activities.

Premack compared the proportions of total time spent in each activity under free conditions with the proportions after he had imposed a 1:1 contingency between the several possible pairs of activities (The free situation, with which the contingent situation is to be compared, is known for some obscure reason as the *paired-baseline* condition). The contingency was like a fixed-ratio 1 schedule, but the units were time rather than responses. For example, the animal might be required to accumulate 10 s of time doing activity A before he would be allowed 10 s of activity B. Activity A here is termed the *operant* or *instrumental* response and B the *contingent response*. Premack (1965) wrote:

The clearest predictions possible were those for Chicko, who in the first procedure [free access to all activities paired baseline] showed three reliably different response probabilities [proportions of time spent]. Indeed, Chicko's protocol made possible three kinds of contingencies: contingent response [the reinforcer] higher than, less than, and, in one case, about equal to the free [operant] response....the outcomes for the three types of contingencies were as follows: (1) contingent response higher than free response produced...an increment in the free response; (2) contingent less probable than free response produced...a decrement...; (3) the one case in which the responses were about equal produced little or no change, increment or decrement. Thus, a monkey that spent 10% of its time playing with L, and 20% playing with P under free conditions, increased its level of L when L, the operant response, had to occur for several seconds in order for P, the contingent response, to be available for several seconds. Evidently, access to a high-probability activity will serve to "reinforce" a low-probability activity. If you watch TV for, say, three hours each day and study for only one hour, we can increase the amount of time you allocate to study by making access to your baseline (three hours) TV allocation contingent upon three hours of study: you may wind up watching TV less, but you will also study more. The reinforcing effect of a high-probability activity on a low-probability one exemplifies the *Premack Principle*, which Premack proposed it as a general principle of reinforcement.

Are there any limits to the principle: Is behavior *probability* really what determines reinforcing effectiveness, for example? Do Premack's results really represent "reinforcement" in the familiar sense? I begin with the last question.

Figure 17



Schedule and time allocation constraints with only two available activities.

Let's look first at the simplest possible theoretical case. The animal can do only two things, and always does one or the other. That is, there are just two mutually exclusive and exhaustive activities. The situation under free conditions is illustrated in Figure 17, which shows the amounts of time devoted to each activity along the x- and y-axes. A given activity distribution is then just a single point whose coordinates are the times taken up by the two activities. The time constraint is shown by the line of negative slope. The term *time constraint* just refers to the fact that if the animal increases the amount of one activity the other must decrease by the same amount. Point B_0, the *free-behavior point,* represents the times taken up by each activity under free conditions. As I've drawn it, the animal spends more of its time in activity A than in activity B. The line of unit slope through the origin represents the schedule constraint imposed by the 1:1

contingent relation between the activities. Remember, i stipulated that x units of operant activity are required for x units of contingent activity, which implies a feedback function of unit slope-the feedback function for a fixed-ratio-one (FR 1) schedule.

The animal's behavior must satisfy both the schedule constraint and the time constraint. The only point that satisfies both is the intersection of the two constraints, point B1 in Figure 17. The arrow indicates the forced change in the proportions of the two activities implied by the shift from B0 to B1: activity A (the contingent, and more frequent, activity) decreases, and activity B (the operant, and less frequent) increases.

Notice two things about Figure 17. First, by making the more-frequent activity contingent on the less-frequent, we cause the less-frequent to increase. But, second, notice that this increase is *not* a "reinforcing" effect of the more probable activity on the less probable. It is not an adaptive response, but merely a forced change, caused by the two constraints. Because the schedule makes operant and contingent activities occur equally often, and because the contingent activity occurs more often than the operant activity under free conditions, the first effect of the 1:1 schedule constraint is artificially to restrict the time devoted to the contingent activity. Because the animal must engage in one activity or the other, restriction of the contingent activity forces an increase in the operant response. This forced change in activity levels is called a *restriction effect.*

It should be obvious that the two constraints, time allocation plus the FR1 schedule, exhaust the degrees of freedom available to the animal in this simple situation. We will see in a moment that to get something more than restriction effects we must have more activities than constraints at least three activities, if we have a schedule constraint in addition to the always present time constraint.

There is a simple experimental test for whether a given increase is just a restriction effect, or something more. The test for a restriction effect is to ask whether the increase in the operant response associated with the imposition of a contingency is greater than, or merely equal to, the increase produced by just restricting the proportion of time the animal can devote to the contingent activity. The effect of restriction is illustrated in Figure 17 by the dashed vertical line through point B1. Under these conditions (two mutually exclusive and exhaustive activities) restriction of the contingent response to the level attained under the contingency condition yields the same increase in the operant response, despite the lack of any contingent relation between the two activities.

The Yoked-Control Method. This test is easy to state in principle, but it is not so obvious how we would go about it in practice. We want to compare the level of the operant response, activity B (e.g., lever pressing) obtained under the contingency with its level when the contingent response, activity A (e.g., wheel running) is artificially kept at the same level as it is under the contingency. One way to do this would be by comparing the two conditions successively. First, in phase 1, we impose the contingency and measure the level of activity A. This might require several daily experimental sessions to get reliable data. Then, in a second phase, we use a timer to restrict access to A (i.e., lock and unlock the running wheel) to the same level it attained in the first phase. For a fair comparison, we also want to control the temporal pattern of A, to keep it the same as it was in the first phase which means that we need to have recorded the time of every occurrence of A in phase 1. This method works quite well, so long as we are sure that the animal's experience in phase 1 has not changed him in some way so that his behavior in phase 2 is different from what it would have been if we had run phase 2 first. For Premack's kind of experiment, this is not usually a problem. But when it is, there is an alternative method which requires us to use two animals at a time. One animal is called the *leader* the other the *follower.* Leader and follower are housed in two separate boxes, both controlled by the same computer. The leader works under the contingency: to get access to A he must engage in B. The follower cannot control access to A; but he gets access to A whenever the leader gets access to A. Thus the follower gets exactly the same temporal distribution of A as the leader, but he gets it independently of his behavior.

We can now compare the level of behavior B for the leader and the follower. If both are equal, we conclude that any elevation in B relative to paired-baseline levels is just a restriction effect. But if the leader shows a higher level of B than the follower, we conclude that we have a real reinforcing effect, termed a *contingent* effect. Contingent effects are the main topic of this chapter.

Both these procedures are known as *yoked-control* procedures. The first is ***within-animal*** the second ***between-animal.*** Both have their limitations. The within-animal method fails if our effects are not completely reversible, i.e., if the effect of either treatment depends on prior experience. The between-animal method fails if (as Premack in fact found) our two animals are not identical: if one animal has a higher free level of A than the other, for example (we can mitigate this problem somewhat by looking at average data from several leader-follower pairs, but this is not always a good solution).

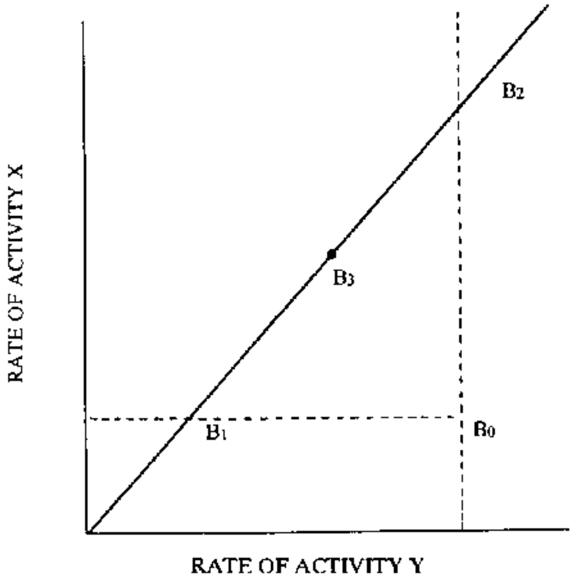
We must rely on our knowledge of the animals and the activities we are studying to judge which is more appropriate in any particular case.

Premack's early experiments lacked the necessary controls to rule out restriction effects. Nevertheless, later work has borne out his principle that a more probable activity will generally reinforce a less probable one on which it is contingent and that this increase is greater than that expected merely from restriction.

Premack thought of reinforcement as being particularly associated with *activities* not with events. For example, with eating rather than the availability of food, or with running rather than the availability of a running wheel. But as we will see, the important thing is just the preferred levels of events, which may be activities but can also be the presentation of stimuli. Consider the case of *punishment*, for example. Receiving electric shock is not an activity. Nevertheless, animals clearly have a preferred rate for such an event, which is zero. With this sole quantitative difference, electric shock can be treated as just the reverse of reinforcement. Requiring an animal to engage in a low-probability act, or experience an event that is generally avoided, for access to one of higher probability has a reinforcing effect on the low-probability act, but it has a punishing effect on the high-probability act. Reinforcement and punishment are thus two sides of the same contingent coin. The main difference is a practical one: commonly used punishers like shock are usually events that can be delivered to the animal without its cooperation. However, this is not an essential difference between reinforcement and punishment. Animals can be punished by forcing them to engage in more of a low-probability activity than they would like, just like the schoolchild required to write 100 "lines", and they can be positively reinforced by hypothalamic brain stimulation which requires as little cooperation as electric shock.

Notice that the imposition of a contingency allows the animal a range of possible options. In Figure 18 the free levels of two mutually exclusive, but not exhaustive, activities are indicated by B, the free-behavior point. A ratio contingency-access to Y depends on performance of X-is indicated by the line through the origin. If it is much more important to the animal to regulate the rate of Y than the rate of X, then we might expect X to increase to the level indicated by point B_2, which allows Y to continue to occur at its baseline rate. But if regulation of X is much more important than regulation of Y, X will remain at *its* baseline rate, forcing a large drop in the rate of Y. In practice, either extreme result is rare and animals usually settle for a compromise solution, such as point B_3, where activity X occurs above its preferred level and activity Y below its preferred level. I will discuss some explanations for this kind of compromise in a moment.

Figure 18



It shows the free levels of two activities mutualy exclusive (activity Y and activity X) but not exhaustive. This levels are indicated by B0; the free behavior point. The line through the origin shows the contingency of Y over X.

In addition to the explicit assumptions we have already discussed, Premack's hypothesis contains an important implicit assumption: that reinforcing effects depend only on *molar* measures such as the total proportion of time devoted to an activity. He was not at all concerned with *molecular* properties, such as the pattern in time of particular activities, or contiguity relations between the operant and contingent responses. Many experiments have shown that continuity is important; a few experiments have also shown that the temporal pattern of an activity, not just its average frequency, can be important. So we may suspect that Premack's theory is at least incomplete. Nevertheless, his emphasis on molar analysis has proven to be a useful approach that has led to several advances in our understanding of behavioral regulation and the functional properties of reinforcement schedules. I take up these topics in a moment, and return later to the molecular analysis of behavioral allocation and choice. But first, I need to say something about the dynamic processes underlying behavioral allocation.

Dynamics of Behavioral Allocation

Over a 30-min period animals like rats or pigeons typically show a varied pattern of activity. A rat, at the appropriate time in its daily cycle, may groom itself, run in a running wheel, drink, and eat during such a time interval. What causes the transition from one behavior to the next? What determines the distribution of activities over this time, i.e., the proportions of time spent in running, eating, etc.? We can get an idea of the possibilities by artificially preventing the animal from doing one activity, such as eating. We can expect two kinds of effects: (a) some change in the distribution of the remaining activities, and (b) a compensating change when the blocked activity is restored.

If access to one activity is blocked, then the time devoted to at least one other activity must increase. A simple stochastic competition model (cf. Staddon, 1988) illustrates the simplest rule for behavior reallocation. In this model, two or more activities compete for expression. The "strength" of each activity depends upon a variable V that varies in random fashion; the activity with the highest V is the one to occur. When there are three activities, whose strengths vary according to the same random process, then each will obviously occur for about one-third the time. Suppose we now block one activity. The situation is reduced to one with two activities, whose strengths vary randomly. Clearly these two will now split the time up equally. This suggests a general rule for behavioral reallocation: the relative proportions of time taken up by the activities remaining after one or more has been blocked will remain constant. For example, suppose we have three activities A, B, and C that take up the following percentages of the total time: A: 50%, B: 25%, C: 25%. If we now block C, this rule implies that A will increase to 67% and B to 33%. This rule is variously termed *Luce's principle* (Luce, 1959, 1977), the *independence of irrelevant alternatives* or the *axiom of choice*.

Experimental results are sometimes in agreement with Luce's principle, but there are some striking exceptions. For example, suppose that the blocked activity is eating, and one of the remaining activities is drinking. If our subjects are rats, eating and drinking are loosely linked under free conditions rats typically drink before and after meals so that omission of eating will usually *reduce* rather than increase the level of drinking, a clear violation of the principle. Luce's principle also assumes a closed set of activities; it cannot account for the appearance of a novel activity for example. Yet without this possibility, little operant learning could ever take place.

What will happen when the blocked activity is restored? Our simple, model says that the previous distribution of activities will be immediately reestablished; there will be no aftereffect of the period of deprivation. Once again, experimental results are very different. If the blocked activity is eating, for example, then after a period of deprivation the animal will eat a lot more than normal: the first meal will be very long (Le Magnen, 1985). This result is typical: a period of deprivation is almost always followed by a compensatory increase when the blocked activity is restored (see Mook, 1987, for an excellent review of recent research on motivational systems like this).

This result is reminscient of a well-established property of motivational systems most graphically described by the hydraulic model of Konrad Lorenz (1952). Lorenz proposed that the tendency to engage in any motivated activity is not constant but *increases* so long as it is *not* occurring (and decreases when the activity occurs). If we are to stick with our general approach in terms of competing V-values, then the fact of compensatory increase implies that the V-value for an activity must increase during a period of deprivation. It seems clear that our stochastic model for behavioral allocation is too simple.

But now we can begin to see the outlines of a dynamic basis for behavioral allocation and regulation: if activities that are *not* occurring increase in strength (a deprivation effect), and activities that *are* occurring decrease in strength (a satiation effect), we have a potentially self-regulating or *homeostatic* system. The system is regulatory because it changes in such a way as to oppose the effects of any imposed change: if an activity cannot occur, its strength (i.e., the tendency for it to occur) increases; if it is forced to occur, its strength decreases. The result in either case is to diminish the effect of the imposed change. Although the details are still obscure, it is clear that a dynamic satiation-deprivation process underlies behavioral regulation.

Behavioral Regulation: The Molar Analysis of Behavior Allocation

Premack only studied schedules in which an amount of time devoted to the operant response gave the animal access to an equal amount of the contingent response (i.e., an FR-1 schedule). But our graphical method of analysis allows us to make predictions about what should happen with other schedules. Look again at Figure 18. Suppose responses X and Y were to be related by a ratio schedule chosen so that its feedback function goes through the point B_0. For example, suppose that under paired-baseline conditions the animal spends 40% of his time doing Y and 10% doing X. This ratio implies that access to Y should be reinforcing for X. But if now we require 1 s of X for access not to 1 s of Y but 4 s of Y, will there be a contingent increase in X? Probably not, because by continuing to do X at his paired-baseline level, the animal continues to get as much access to Y as he needs to maintain *its* paired-baseline level.

What about the converse possibility: suppose we require the animal to engage in response Y (high probability) for access to response X (low probability), but make the terms of the exchange so unfavorable that the animal can only maintain X by increasing the level of Y? For example, we require 5 units of Y for access to 1 of X. Even though X is lower probability than Y (so should not be able to reinforce Y, according to Premack), our analysis suggests that Y should in fact be reinforced under these conditions, because the paired-baseline ratio of the two is 4:1, which is less than the 5:1 imposed by our schedule. Experiments have shown that under these conditions the level of Y will usually be increased. These results led William Timberlake and James Allison (1974) to propose a generalized version of the Premack principle that they called *response deprivation*: they proposed that a contingency between one activity and another should change the level of the operant response if the animal cannot attain his paired-baseline levels without changing the level of the operant response. In terms of our diagram, response deprivation amounts to the assertion that the operant response will increase if the schedule constraint (feedback function) lies above the free behavior point (as in Figure 18); and will decrease if the schedule constraint lies below the free-behavior point.

Response deprivation is a regulatory hypothesis, since the core idea is that animals act so as to maintain a certain distribution of activities: If the imposition of a contingency means that any activity is pushed below the animal's preferred (paired-baseline) level, other activities will be reallocated so as to reduce the discrepancy. If we measure the amount of drinking and bar pressing during a baseline session, for example, we are likely to find that animals drink at a particular rate and bar press very little. If we make drinking at the baseline rate contingent upon bar pressing above baseline rate, bar pressing is likely to increase. But how much will it increase? And what will be the effect on other activities, such as wheel running and sleeping? Response deprivation is only qualitative: it says when there will be a change, and what its direction will be, but it does not say *how much* change there will be. I turn now to regulatory theories that attempt to answer these questions.

Behavioral Conservation. Behavioral conservation is a model of behavioral regulation proposed by James Allison (1980). Conservation theory assumes that animals attempt to conserve some property of the set of activities. Allison has tentatively proposed energy expenditure as the conserved dimension (but I will have an alternative suggestion). If energy is conserved, then the energy expended on all activities in the paired-baseline session should be same as the energy expended when access to water is contingent upon bar pressing. Allison formalized the conservation model by the following equation

$$kND+NP = kO_d+O_p, \tag{1}$$

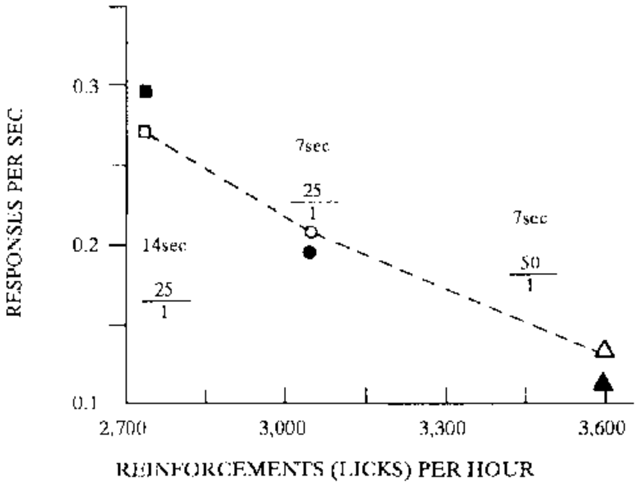
where O_d and O_p represent the amounts of drinking and lever pressing that occur during the paired-baseline session and D and P are the amounts "traded" during each cycle of the contingency session (for example, 5 s of drinking for 10 s of lever pressing) and N is the number of cycles. Notice that the quantity ND is just the total amount of D in the contingency session, and NP the total amount of P. Notice also that the right-hand side of eq. 1 is constant, since k is constant and so are the baseline levels of the two activities. Thus, eq. 1 says that the total amounts of D and P during the contingency session are equal to the total amounts during the paired baseline, with D given a weight of k relative to P in both cases. Constant k represents how energetic D is relative to P. For example, suppose each instance of D (lick) takes half the energy required for each instance of P (lever press); then conservation says that 2 x number of licks+number of lever presses will be constant over any fixed time period.

Allison's notation is cumbersome. It is easier to see what is going on if we just deal in total amounts of activity and ignore the number of cycles, N, which is irrelevant to the theory. With this simplification, eq. 1 reduces to $ky+x=$ constant, where y is the total amount of P (i.e., NP in the contingency session, O_p in the baseline session) and x is the total amount of D (i.e. ND in the contingency session and O_d in the baseline session). We can therefore rewrite the conservation model as a linear relation between the total amount of drinking, x, and lever pressing, y:

$$y = K - x/k, \tag{2}$$

where K is a constant proportional to session length, and k is a constant that represents the relative importance of x and y. Equation 2 says that the amount of Y, the operant response, is related to the amount of X, the contingent response, by a straight line with negative slope -1/k.

Figure 19



Responses per second of rats exposed to a VI 7 seconds or VI 14 seconds, with 25 or 50 lickings to a bottle with water as reinforcer. The broken line shows the fit of the data.

Allison has obtained many experimental results that support this linear relation. An example is shown in Figure 19. In this experiment, rats were exposed to either VI 7 or VI 14 second schedules where one lever press produced access to either 25 or 50 licks of water. As you can see, no matter how much water the animals received for each effective lever press, the total number of lever presses was related to the total number of licks by the same straight line of negative slope. The fit to data looks excellent, and Allison has presented many other equally compelling results; yet conservation theory is not generally accepted as a valid model for reinforcement. What might be wrong?

An obvious problem is that conservation theory says nothing about the *schedule* relating the two behaviors, X and Y. It says that the relation between x (the rate of X) and y is linear, no matter what the schedule relating the two. Yet, as we will see, there are numerous experimental results showing that the schedule makes a great deal of difference. For example, animals always respond faster on ratio than on interval schedules, even if both schedules yield the same rate of reinforcement. If, in a more elaborate experiment, we measure the rate of lever pressing under a range of ratio schedules and plot it against the rate of food delivery obtained on each schedule we usually get a different function than if we plot lever pressing against food delivery over a range of variable-interval schedules.[78] Conservation theory must predict the same relation for both. Yet Allison has presented many data, such as those in Figure 19, that seem to support conservation and show the same function for interval and ratio schedules. How can this contradiction be resolved?

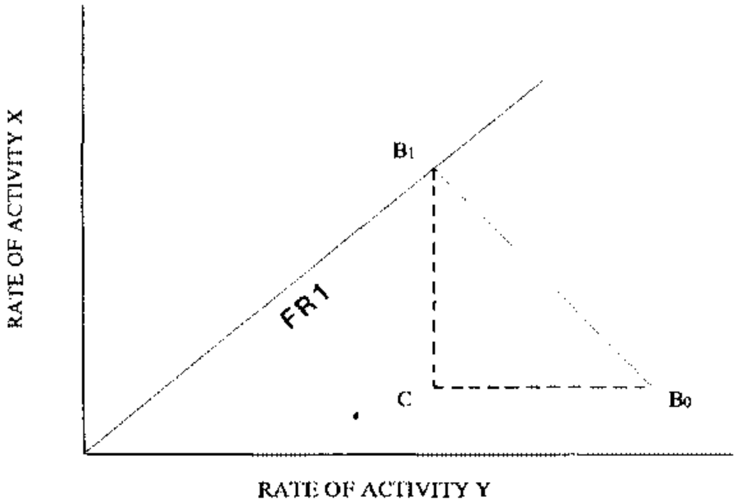
The conditions under which the data in Figure 18 were gathered provide a clue to the answer. The variable-interval schedules Allison used to collect these data are very short-7-15 s compared to typical VI values, which are usually in the range of minutes, and the reinforcement 5 or 50 licks was rather long. His rats were undoubtedly pressing their levers at a very high rate on such rich schedules, and perhaps had little time for anything but drinking and pressing. But if his animals were in effect engaging in two mutually exclusive *and exhaustive* activities then the situation is a familiar one. It is just the case described in Figure 17, in which the relation between the two behaviors is completely determined by the time constraint. When the time constraint is dominant, then the schedule indeed makes no difference and X and Y will be related by a straight line with negative slope, just as Allison has reported.

We can conclude, therefore, that the conservation model probably amounts to no more than a time constraint: something is indeed conserved, but it may be nothing more than the total time available. The constant k then represents the relative durations of the two activities. If a lick (X) takes only half the time of a lever press (Y), for example, the conservation relation has to be $2x+y$ = constant, where x and y are the rates (number per minute) of the two activities. The conservation principle therefore tells us nothing about the nature of reinforcement, but only something about the limitations on our methods for studying it. I turn now to more sophisticated regulatory models that can deal with the different molar patterns generated by different reinforcement schedules.

Optimal Policy on Ratio and Interval Schedules

Look again at Figure 18. Recall that if a ratio contingency is imposed that forces behavior away from the free-behavior point, B_0, animals typically settle for the sort of compromise represented by point B_3: the contingent response, Y, decreases somewhat (but not as much as if X had remained at its paired-baseline level, i.e., not to B_1), and the operant response, X, increases (but not as much as necessary to maintain the level of Y at *its* paired-baseline level, i.e., not to B_2). I first diagrammed reinforcement schedules in this way in 1976 and the diagram immediately suggested a simple hypothesis: Suppose that the animals just get *as close as possible* to the free- behavior point, which all theories acknowledge as the animal's preferred state (economists call points like this *bliss* points). This prediction is illustrated in Figure 20: it just amounts to dropping a perpendicular from B_0 to B_1 on the straight-line ratio-schedule feedback function. B_1 corresponds to the compromise position B_3 in Figure 18. The *minimum-distance* (MD) prediction (Staddon, 1979) always falls somewhere in between the two extreme possibilities. Now we can test the model by seeing if an FR 1 produces the predicted increase in response X.

Figure 20



Relations between the schedule constraint (line through the origin) and the distribution of activities under paired-baseline conditions (point B0). In most of the discussion we assume that the schedule is *reciprocal*, e.g., on FR 1, one unit of Y is required for access to one unit of X, which must be taken up before access to Y is possible again. On standard reinforcement schedules, the constraint is usually only one-way: there is a requirement to engage in activity X (e.g., lever pressing) for access to activity Y (eating), but X is always available. Since Y is always highly preferred, the reciprocal contingency is assumed to be unnecessary.

The first results were discouraging. If response X is key pecking and response Y is eating, by a pigeon at 80% of its normal weight, the predicted increase in the level of key pecking is much too small. But the reasons are not far to seek. The first problem is one of *units*. In conventional operant conditioning experiments key pecking is measured as pecks per unit time and reinforcement rate as food deliveries per unit time, but there is nothing that makes these two equivalent: a peck does not take up as much time as a food reinforcement (which is typically three or four seconds access to a grain hopper). At the very least, the predicted number of pecks needs to be increased to reflect their shorter duration. But even with this correction in the vertical scale, the predicted increase in pecking falls far short of what is observed. On an FR 2 schedule, for example, a hungry pigeon may peck almost twice as fast as on an FR 1, yet the uncorrected MD model shows him responding at about the same rate. Evidently something is still missing.

The Premack principle equates the reinforcing value of an activity with its probability-how much time the animal spends in the activity. Yet a little reflection suggests that this cannot be the whole story. The comic strip cat Garfield spends most of *his* time resting, but we know that *food* is what he finds reinforcing. Under many conditions probability is correlated with reinforcing strength, but anyone can think of counter examples-sexual activity, for example, which is relatively infrequent yet highly reinforcing. The important difference between rest and eating has to do not eith the free levels of each, but with the degree to which those levels are *regulated*. Garfield can tolerate interruptions of his rest a great deal better than curtailment of his food supply-and you are probably the same. How can we incorporate this property into the MD model?

Look at the triangle B_1B_0C in Figure 20. It just makes visible the separate deviations of behaviors X and Y from their free levels forced by the FR schedule: distance CB_0 is the deviation in the level of Y (eating), distance CB_1 is the deviation in the level of X (key pecking). With an FR 1schedule, these deviations are equal. Now which deviation is more upsetting to our hungry pigeon, do you suppose? Obviously CB_0, the shortfall in eating rate, is likely to be a great deal more important than CB_1, the surplus of keypecking. It is relatively easy to modify the MD model to take care of the different costs of deviation (CoD's) in the levels of the two behaviors. To see how this is done, I need to explain two new concepts: the concept of cost, and the concept of an objective function.

Cost and Objective Functions. The conservation model assumes that there is some property of the set of activities that is conserved (maintained constant) in any closed situation. It turns out that the quantity conserved is probably the total time taken up, but it might just as well have been the total energy expended or something even more complicated. The MD model takes an important step beyond this and assumes that some quantity associated with the set of activities is *minimized*. This represents a great advance, because now the properties of the reinforcement schedule enter in as constraints on the minimization so that minimization (optimality) models do indeed make different predictions about performance on different schedules.

Let's look at the general method for doing an *optimality analysis* (as this is called). The method involves four steps:

(a) First decide on what property (or properties) of behavior is free to vary this is termed the *dependent variable* (another term is *strategy set*). In the present case, the dependent variable is just rate of key pecking.

(b) Then decide what is to be minimized (or maximized). This quantity is termed the *cost function* (or value function). The choice of cost function is critical and also difficult, because there are always very many possibilities. I will begin with the simplest possibility, namely interfood interval, and then look at the effect of modifying it.

(c) Then identify the *constraints* in the situation. There are just two here: time (the time taken by all mutually exclusive and exhaustive activities must add up to the session length), and the fixed-ratio schedule (which constrains the relation between peck rate and food rate). I will ignore the time constraint in this elementary discussion because its effects are usually minimal. We will focus on the schedule as our constraint.

(d) Then derive a quantitative expression for the total cost as a function of the level of the dependent variable, subject to the constraints. This sounds complicated, but in the present case it just means deriving an expression for the average interfood interval (cost) as a function of peck rate (dependent variable) when food rate depends on peck rate according to a fixed ratio schedule (constraint). This expression, which relates the behavior to the total costs, subject to the operative constraints, is termed the *objective function* because it defines the quantity that must be minimized.

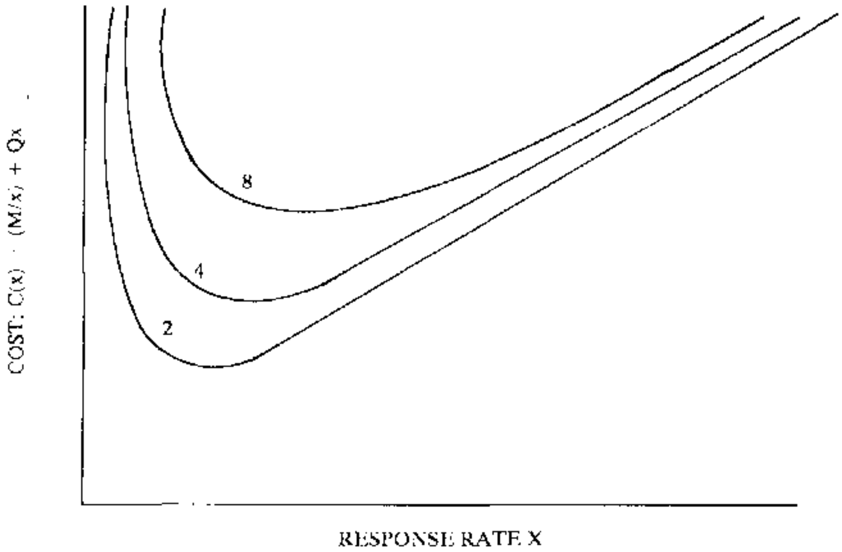
On an FR schedule, the time interval between food deliveries, which we will call I(x) (to indicate that it is a function of x), depends on response rate, x, according to the following relation

$$I(x) = M/x, \tag{3}$$

where M is the number of responses in the ratio: if the animal responds 10 times a minute and gets food on an FR 5, he will get food at 30 s intervals. Equation 3 is the objective function for interfood-interval minimization on FR schedules.

It doesn't take calculus to see that to minimize this objective function, x should be as large as possible: the animal should respond as fast as he can. This is also the commonsense conclusion: if getting food as fast as possible is what is important, then on a ratio schedule you need to respond as fast as you can. This conclusion is not very helpful because it gives the same answer for any ratio value: no matter what the value of M, the optimal policy is to go flat out. But we know that animals go at different rates on different ratio schedules. So what's wrong with our analysis?

Figure 21



The effect of response rate on cost, where cost is defined as interfood interval plus response rate, weighted by a factor Q: $C(x) = M/x + Qx$, *where M is the ratio value. The three curves are for ratio values of 2, 4 and 8, and Q = 0.5.*

One answer might be that we have ignored the time constraint, but since that affects every activity it doesn't actually alter anything. What about the cost function? An obvious possibility is that we have neglected the cost of key pecking: food delay may be costly, but surely key pecking, especially at the high rate characteristic of ratio schedules (the prediction about this is correct, at least) is also costly. So let's add the assumption that pecking incurs a cost proportional to its rate. This just means adding a term Qx to the objective function, where Q represents the cost of each peck per unit time and x is peck rate. Our cost function now has two components: time (interfood interval) and response, scaled in time-cost units (Q). Thus, the new delay-plus-response-cost objective function, C(x), is:

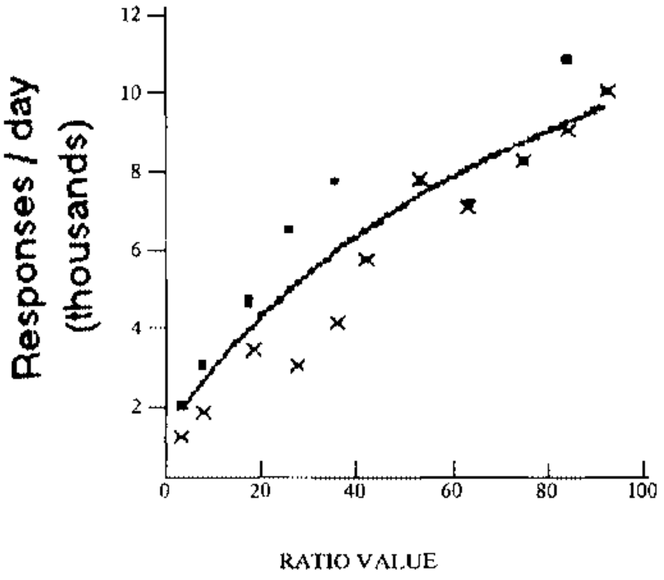
$$C(x) = M/x + Qx. \tag{4}$$

I'll call this the *response-cost* (RC) model. Now we can easily find the value of x for which C(x) is a minimum. The result is shown graphically in Figure 21, which shows Equation 4 plotted for three values of the ratio value, M: 2, 4 and 8. Notice two things about these curves: (a) The minimum cost occurs at a finite response rate. When each response is costly it doesn't pay to go flat out all the time. (b) The response rate that minimizes cost *increases* with ratio value. (The reason is that C(x) is made up of two components: M/x,, which declines as x increases, and Qx, which increases with x. The point where the influence of the Qx component becomes dominant shifts to the right as M increases.)

Both these features are encouraging. Response rate on ratio schedules is high, but usually less than "flat out". And empirical results from rats, pigeons and several other species all show that over most of the typical range of ratio values, response rate does indeed increase with ratio value.

Some typical experimental results with ratio schedules are shown in Figure 22. The data are from a group of Guinea pigs each pressing a lever for water reinforcement. The animals had access to the lever all day and obtained all their water ration via the schedule. The figure shows number of responses per day plotted against ratio value. As you can see, response rate increases with ratio value. The heavy curved line is the prediction of our response-cost (RC) optimality model. It fits the general form of the data quite well, even though it is not in fact a very good model. As we saw with conservation theory, the fact that a model can generate a simple curve that fits data well is a rather weak kind of support for the model. It is much more important that a model provide an accurate picture of a *pattern* of experimental results. The response-cost model does much better than conservation theory in this respect, but not quite as well as the minimum-distance model or more elaborate economic models.

Figure 22



Daily number of lever presses on an ascending series of ratio schedules of water reinforcement for a group of Guinea pigs. Filled squares: group that got 10 s access to water as reinforcement; x's: group that got 20 access (replotted from Hirsch & Collier, 1974, Fig. 1). The heavy curved line is the optimal policy predicted from a response-cost model: x =sqrt(M/Q), where M is the ratio value and Q the cost of each response.

Now that you have a good understanding of the idea of an objective function, we can return to the problems of the minimum-distance model. Look again at Figure 20. Recall the basic assumption: that the animal minimizes the deviation B_0B_1. B_0B_1 is just the square root of the sum of the squares of the two other sides of the triangle, CB_0 and CB_1, i.e., $(x_1-x_0)^2+(y_1-y_0)^2$, where x_1 and y_1 are the coordinates of point B_1 and x_0 and y_0 are the coordinates of B_0. If we minimize the square root of something we also minimize the thing itself. Thus, the objective function for the simple minimum-distance model is just

$$C(x)=(x_0-x_1)^2+(y_0-y_1)^2.$$

If we now add a parameter to reflect the greater importance of regulating activity Y (say, if Y is eating and X key pecking) we arrive at the formula

$$C(x)=(x_0-x_1)^2+c(y_0.y_1)^2, \tag{5}$$

which is the objective function for the MD model. Parameter c is termed the *cost-of-deviation* (CoD) for activity Y. Notice that there are three features about the MD objective function that are different from the simple RC model. First, it is a two parameter model, since each activity is characterized by two parameters, its paired-baseline level (e.g., y_0) and its cost-of-deviation (e.g., c). The RC model has only one parameter, Q, which represents the relative importance of response cost and reward delay. Second, the MD model is based on the rates of activities, rather than their delays. And third, it is nonlinear in the sense that it assumes that the cost of a given deviation increases as the *square* of the deviation, so that doubling the deviation increases the cost by a factor of four. The latter makes good intuitive sense. A drop in feeding rate of 10 gm/hour will obviously be more costly to the inhabitant of a prison camp, starving on a subsistence diet, than to a well-fed suburbanite. The further the feeding rate from the bliss point, the more costly additional deviations become. This idea of positively accelerated cost (which is equivalent to negatively accelerated value) has interesting implications for choice behavior, as we will see in a moment.

The objective function with CoD parameter solves the problems with the MD model that we identified earlier. If the CoD parameter, c, is much larger than unity, an MD animal will freely increase the rate at which he makes low CoD response X so as to maintain approximately constant the level of high-CoD response Y. If c is high enough, an MD animal will almost double his response rate when we increase the FR value from one to two, for example, thus maintaining reinforcement rate almost constant. With these amendments, the MD model makes tolerable predictions of the empirical relations between response and reinforcement rates on variable interval, variable-ratio and many other reinforcement schedules. In most cases, the predicted relation is an inverted-U: response rate is low at very low and very high reinforcement rates, highest at intermediate rates. The function for ratio schedules is tipped to the right and is everywhere above the function for variable-interval, which is tipped to the right.

The MD model is not particularly simple algebraically, nor does it make strikingly better predictions than some other optimality and economic models. But it is important historically as one of the first attempts to show how well-known molar patterns of behavior on reinforcement schedules might be explained by a unifying optimality analysis. It also brought out an important difference between strong reinforcers, like food for a hungry pigeon, and weak reinforcers, like the opportunity to play with a plunger for a Cebus monkey: feeding is a highly regulated activity, whereas play is not. In Premack's original view, everything of importance about an activity is contained in its free, paired-baseline level: point B_0 in Figure 4, coordinates x_0 and y_0 in Equation 11.5 (the so-called bliss point). The MD model showed that activities differ in more than one way: in their free levels, yes, as Premack pointed out; but also in the degree to which they are regulated, indicated by a parameter that represents the different costs of deviation from the bliss point. The higher the CoD parameter, the better an activity will be regulated the harder the animal will work to maintain it at its free level. Strong reinforcers like food may or may not have high operant levels; they certainly have high costs-of-deviation. Weak reinforcers, like the opportunity to run in a wheel, may or may not have low operant levels; they will certainly have low CoD parameters.

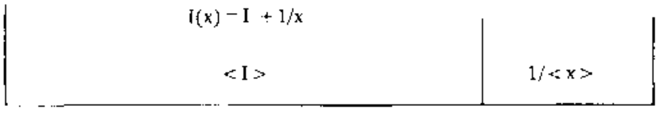
The details of the MD model may now be largely of historical interest, but it embodies two features that are likely to be essential to any adequate molar theory of behavioral regulation: (a) at least two parameters for each activity; and (b) the idea that a comprehensive optimality model must assume a nonlinear cost function.

Now we are in a position to see in detail how an optimality analysis can be used to make predictions about how animals should behave on different reinforcement schedules. But because the MD model is algebraically difficult we will use our simple response-cost model which is a bit less accurate than the MD model, but not as different as it might appear and much easier to understand. You have already seen how to derive the prediction that on ratio schedules response rate should be directly related to ratio value (see Figure 22). Now let's tackle a more difficult case: variable-interval schedules.

Molar Optimal Policy on Variable-Interval Schedules. I will follow the four steps described above. Luckily, everything is the same as for ratio schedules, except the schedule constraint. How does interfood interval depend on response rate on VI schedules? VI schedules are much more complicated than ratio schedules, surely? Not really. Figure 23 shows that on a VI schedule the average interfood interval is the sum of two delays: the average delay set by the VI timer, <I>, plus the average delay between setup and the reinforced response, which is determined by the animal's average response rate, if response rate is random in time, this average delay is equal to the reciprocal of the average response rate, i.e., 1/<x>. Putting these two together we arrive at the following expression for the average interfood interval:

$$I(x) = 1 + 1/x. \tag{6}$$

Figure 23



Components of the average interfood interval on variable-interval schedules

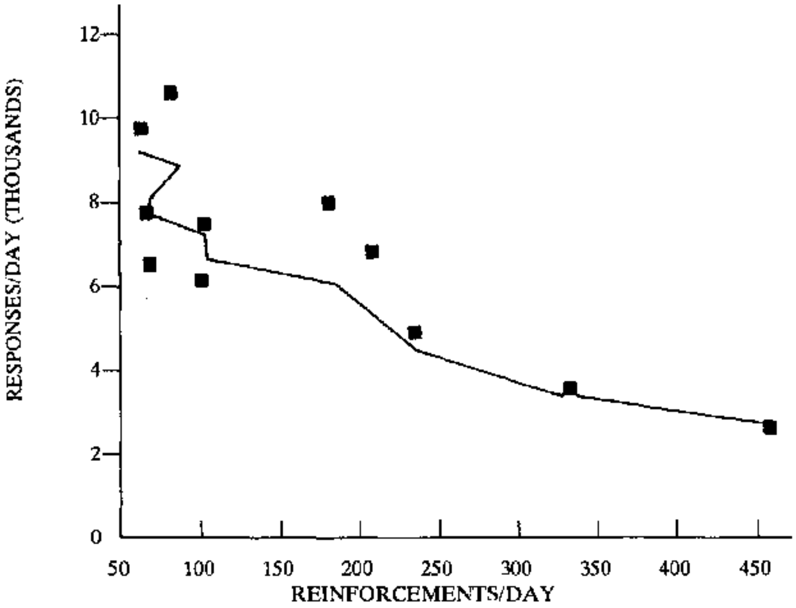
To find the optimal policy, we can now just plug this expression into Equation 4 in place of M/x, the expression for interfood interval on ratio schedules. The result is

$$C(x) = I + 1/x + Qx. \tag{7}$$

We can then go through the same graphical analysis as we did before, but to cut a long story short the conclusion is exceedingly simple. To minimize this cost function, our animal should respond at a constant rate: $x = 1(1/Q)$ (as opposed to $1(M/Q)$ for the ratio schedule). How do the molar data from VI and FR schedules compare with these predictions?

The relevant data for pigeons, key-pecking for food reinforcement on VI schedules, are well known: the rate of key pecking is a negatively accelerated function of obtained reinforcement rate (Catania & Reynolds, 1968). But the function is indeed approximately constant over quite a wide range as the RC model implies it should be. The function is a very reliable finding for both rats and pigeons on VI schedules (when rates are measured so as to exclude the time spent eating from the time denominator). The data from ratio schedules are somewhat more variable, depending on exactly how they are obtained whether in a so-called *closed economy*, in which the animal is able to respond throughout the day and must get all his food via the schedules, or in an *open economy*, with short, daily experimental sessions and supplemental feeding after the session to make up any nutritional shortfall (Hursh, 1984). Figure 22 is typical of closed-economy data, which invariably show that response rate is directly related to ratio value and, usually, inversely related to the rate of reinforcement actually obtained. The data from Figure 22 are replotted in this way response rate vs. obtained food rate in Figure 24. Data from open-economy experiments are often of this form, but sometimes also turn down at very low reinforcement rates (corresponding to very high ratio values; cf. Timberlake & Peden, 1987). As you can see in Figures 22 and 24, the RC prediction is quite good.

Figure 24



Data from Figure 6 replotted as response rate vs. rate of reinforcement obtained, for the 10-s access condition. The solid line is the optimal policy according to the response-cost model, as before.

Let's compare directly the predictions of the RC model for VI and FR (or VR) schedules. On FR, response rate is determined according to the relation x= 1(M/Q), where M is ratio value and Q is response cost. The relation for VI is very similar: x = 1(1/Q); in other words, the RC model says that the animal should treat a VI schedule just like FR 1. But notice what these two equations predict about the *relative* response rates on VI and VR: for any ratio value greater than one, response rate should always be higher on the ratio. This is one of the oldest and most reliable findings about ratio and interval schedules. Several careful comparisons have been made (e.g., Catania, Matthews, Silverman & Yohalem, 1977; Lea & Tarpy, 1982). In the yoked-control study by Catania et al., for example, the key-peck rates of a leader and a follower pigeon were compared. The leader produced food reinforcement according to a variable-ratio schedule. Reinforcement for the leader constituted schedule "setup" for the follower, who collected the reinforcement with his next peck. Thus, the leader received food on a VR schedule, the follower on a VI, approximately matched to the VR both for temporal pattern of food deliveries and their average rate. Catania et al. found that the VR leader birds always developed faster key-peck rates than their VI followers. Almost any optimality model that includes the time between reinforcements in its cost function will predict this result.

Thus, both the MD and RC optimality models predict the most obvious difference between interval and ratio schedules: the fact that animals always respond faster on ratio schedules. They do so because the different feedback functions enter in to the objective functions as different constraints. I will be more explicit about how these different constraints give rise to different predictions when i discuss marginal utility in the next section. Both optimality models also do a reasonable job of predicting other molar properties, such as the way response rate increases with ratio value, and the approximate constancy of response rate on variable-interval schedules. You may have noticed, however, that there are some inconsistencies in the data: animals in so-called open economies behave rather differently than animals in closed economies, for example. I return to these issues later when i discuss optimality in the context of behavioral economics.

Behavioral Allocation: Conclusion

What have we learned from this account of behavioral allocation? The first step was taken by Premack, who took to heart Thorndike's conclusion that reinforcement must be defined in terms of the organism's own behavior. Premack saw that at the instant an organism initiates an act, that act is by definition the most preferred. If reinforcement is defined by preference (as Thorndike believed) then this act at that moment must be the most reinforcing activity in the organism's repertoire. How might this insight be used to redefine the concept of reinforcement? We can reconstruct Premack's reasoning as follows: At the instant of choice, the act that occurs is the *most probable* act. Probability cannot be measured instant by instant, but we can get an idea of what is most probable by looking at the proportion of time taken up by different activities. Perhaps the most frequent act is also the most preferred, hence the most reinforcing? And so it proved, in the situations Premack and his successors chose to study.

But now that the links in this chain of argument are exposed, you can probably see its weaknesses. The argument blurs the distinction between *molecular* (the instant of choice) and *molar* (the frequency of an activity). It also blurs the distinction between behavioral dynamics and what we might call behavioral *statics*. The momentary-probability definition of reinforcement is about dynamics and the molecular structure of behavior. The activity-frequency definition is about static, molar properties of behavior. Once we agree to look at averages, i.e., at molar behavior, we at once encounter the problem of the averaging *window*, the time period over which we are to count activities. The window size makes a great deal of difference. Premack's intuition about the reinforcing effect of the highest-probability act implicitly assumed a time window short enough to just encompass the instant of choice. But to test his theory, he had to average much longer time intervals. The fact that his predictions were generally confirmed is fortunate and implies that preferences may change relatively slowly. But as we have seen, to extend his theory to "strong" reinforcers like food and water requires additional assumptions.

The real meaning of the molar-molecular distinction will not be properly clarified until we have a full, dynamic understanding of behavioral allocation. But historically the weak points in Premack's original position were shored up by adding assumptions to the molar theory: first the response-deprivation hypothesis, that allowed for schedules other than 1:1; and then the minimum-distance model, which added a parameter, the cost-of-deviation, that took account of differences between overall frequency and momentary priority. MD models allow for activities that are frequent, but not urgent, and the converse; whereas for the Premack principle, and response deprivation, frequency and urgency are the same thing.

Premack set the study of behavioral allocation on a molar path. With a few isolated exceptions, such as the neglected work of David Birch (e.g., Atkinson & Birch, 1970), advances were all in the direction of increasingly refined molar models. The MD analysis, and the work of Howard Rachlin and his associates (1978; Rachlin, Battalio, Kagel, & Green, 1981) on economic models that was going on at the same time, showed the possibilities of molar optimality analysis as a way of understanding the varied effects of different reinforcement schedules on behavior. I have just described a very simple optimality analysis of interval and ratio schedules, and although the fit to data is not exact in any particular case, the analysis should impress by the variety of predictions derivable from an exceedingly simple model and their approximate correctness. Simple optimality analyses like this are the pigeon equivalent of the economists' *rational man*: no real man, woman or pigeon behaves like this, but the model provides a rough sketch of reality that is close enough to be useful. When a mechanistic model comes along, you may be sure that it will approximate the predictions of the optimality model under very many conditions just as our ideal pigeon in the discussion of concurrent chain schedules frequently approximated the optimal policy.

There are still a number of situations for which we have no adequate mechanistic account: choice between rewards of different types or occurring in variable amounts, for example. These are the province of behavioral economics, to which i now turn.

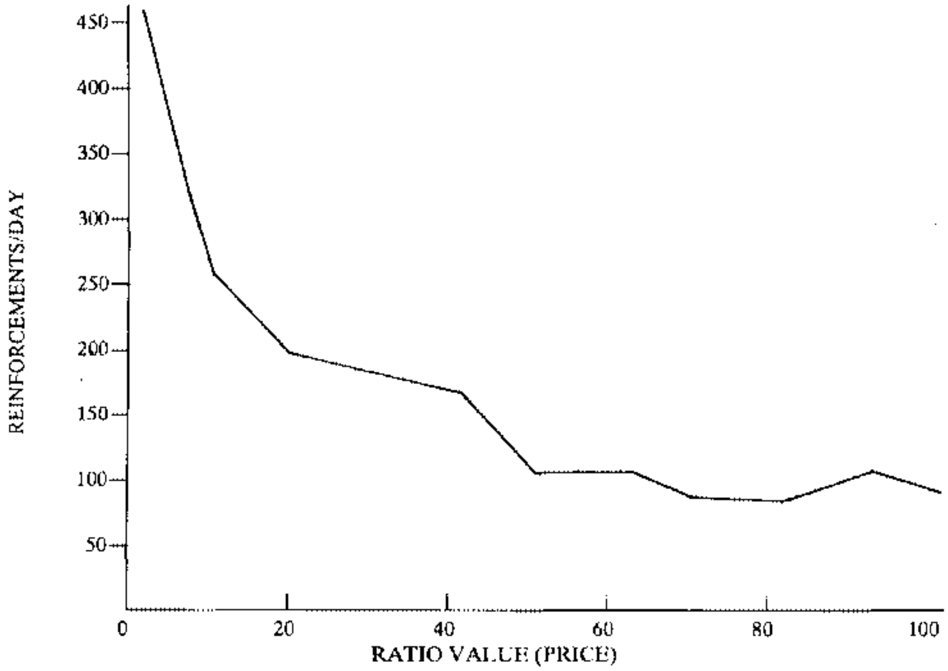
Behavioral economics

A question at the core of economics is, what is the source of value? Why is it, for example, that a 400 carat diamond, too large for a jewel, too small for a paperweight, may fetch a price of millions of dollars, whereas bread, without which man cannot live, is priced cheap? One theory had to do with what we would today call the *inputs* necessary to produce a commodity, labor and time. Bread is cheap, the labor-time theory has it, because it is easy to produce; large diamonds are dear because they are hard to find. But then so are four-leaf clovers, and they have no market price at all. Scottish scholar Adam Smith, the first modern economist, in his great work *An Inquiry into the Nature and Causes of the Wealth of Nations* (1776), solved this problem. We cannot know anything about value, he argued, we can only know about price. (Oscar Wilde once remarked that a cynic is a man who knows the price of everything and the value of nothing; perhaps his cynic was a Smithian economist). About price, Smith came to the same conclusion that we have come to about *reinforcement*: in the end, it is defined by people's behavior we will soon come to the same conclusion about value. He reasoned as follows:

> The market price of every particular commodity is regulated by the proportion between the quantity which is actually brought to market, and the demand of those who are willing to pay the...price....When the quantity of any commodity...falls short of the effectual demand....the market price will rise... (1776/1976, p. 73).

In modern terms, the price of a commodity is explained as an equilibrium between two curves: a rising curve, which says how much will be produced at a given price, the *supply* curve; and a falling curve, which says how much will be bought at a given price, the *demand* curve. The higher the price consumers are willing to pay, the more items will be produced; conversely, the higher the price rises, the fewer the customers willing to pay it. The price at the point the two curves cross is the *market* clearing price, the price at which all who wish to sell are able to sell, and all who wish to buy are able to buy. The demand curve is at the heart of *microeconomics*, the study of how the behavior of individual economic units determines the state of the economy as a whole.

Figure 25



Molar ratio-schedule performance plotted as a demand curve. The data from Figure 6, for the 10-s condition, are plotted as reinforcements obtained per day vs. "price", i.e., ratio value.

About 10 years ago, several researchers pointed out that behavioral data from animals working on ratio reinforcement schedules conform to the law of demand: as the "price" (work requirement) increases, the amount "bought" falls (cf. Lea, 1978). Figure 25 shows an example, which is replotted from a data set we have already seen: it shows the number of reinforcements per day obtained ("bought") under different ratio schedules ("prices") in the Hirsch and Collier (1974) experiment. The curve is typical: as the ratio value increases, the amount of food obtained falls. The "demand law", that as price increases effective demand falls, is almost as general in the world of reinforcement schedules as in the economic world. (Even in economics there are a few exceptions, however: so-called *Giffen goods*[1], which are bought more, not less, as their price increases. Staple foods and prestige items are sometimes Giffen goods).

Demand curves illustrate the essentially regulatory character of responding on ratio schedules, although not perhaps as directly as the response rate vs. reinforcement rate plot shown in Figure 18. A horizontal demand curve indicates a perfectly regulated activity: no matter what the cost, the amount bought holds constant. If we had to pay for the oxygen we breathe, it would show highly *inelastic* demand like this. A downward-sloping demand curve indicates imperfect regulation, *elastic* demand. Complete absence of regulation corresponds to a hyperbolic curve.

There is another economic analogy that has been applied to data like this. A plot of response rate against ratio value (e.g., Figure 22) can be compared to what is termed the *labor-supply* curve, which shows how the amount of labor supplied (response rate) depends on the wage rate (ratio value). The typical labor-supply curve is *backward-bending* (by the convention in economics, the x and y axes are reversed, and the x axis reads from right-to-left rather than left-to-right; in our coordinates, the curve would be an inverted U). The idea is that when the wage rate is low, people are not willing to work much-the supply of labor (by an individual or a group) will be low. As wage rate rises, the amount of labor supplied rises to a maximum; but when the wage rate is very high, labor is withdrawn, because now people have enough money to want to take time off to spend what they have earned (The labor-supply curve is closely related to another curve that became notorious as part of the "voodoo economics" of which the U. S. Reagan administration was accused during its early years, the *Laffer curve* see Box 2).

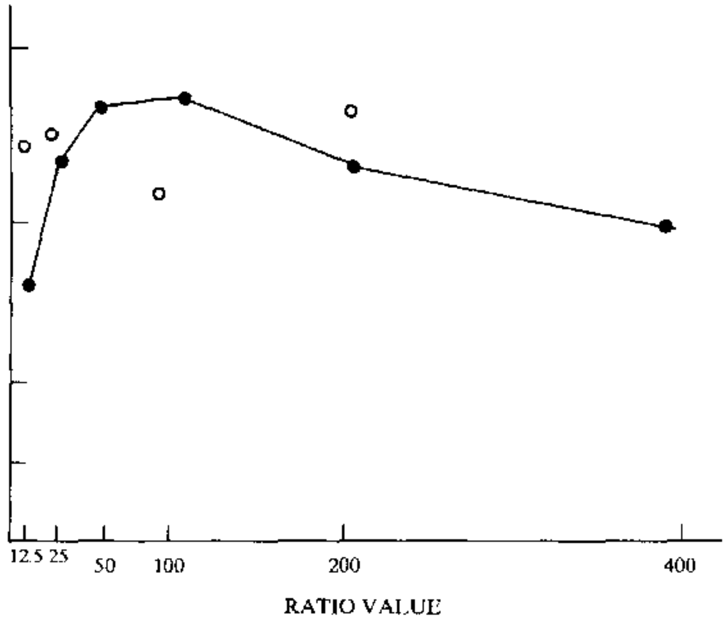
A reinforcement schedule example that fits the labor-supply analogy is shown in Figure 26. It shows data from a single pigeon pecking a response key for food reinforcement on a wide range of random-ratio schedules, ranging from 12.5 to 400. The data are typical of results from open-economy experiments in which the animals are given many sessions of exposure to each ratio value. The functions are inverted-U shaped: the animals respond at a low rate when the ratio value is very high or very low, and at a high rate at intermediate ratio values. Notice that these data are

---

[1] After Robert Giffen (1837-1910), British statistician and economist.

different from the monotonically increasing response-rate vs. ratio-value function in Figure 26.

Figure 26



Molar ratio-schedule performance plotted as a labor-supply curve. Response rate vs. random-ratio value for a single pigeon exposed for many experimental sessions to each of a series of random-ratio schedules (from Green, Kagel, & Battalio, 1982, Figure 18.3).

These two sets of data obviously raise two questions: (a) Which economic analogy is valid, demand curve, or labor-supply? And (b) why are the data in Figure 26 different in form from the data in Figure 22. I'll answer the second question first. As always when attempting to understand behavioral data, the first thing to look at is the procedure: how, exactly, were these data obtained? The demand-curve data (Figures 22 and 25) were obtained from a group of Guinea pigs, in a closed economy (24-hour access to the reinforcer) and water reinforcement. The labor-supply data (Figure 26) were obtained from a pigeon working for food in an open economy (short, daily sessions, supplemented by extra food after the experimental session if necessary). We obviously have a lot to choose from in deciding which of these many procedural differences species, type of reinforcer, length of session, type of economy, individual vs. group was responsible for the different results obtained. In fact, other data pinpoint the type of economy as the indirect cause for the different results *indirect* because in a closed economy there is a limit to the ratio value that can be tried: if the ratio is so high that the animal cannot obtain his minimum daily requirement of water (or whatever the reinforcer is) in the experiment, then he, and the experiment, will soon come to an end. This is not a problem for the open-economy experiment, because any shortfall during the experimental session can always be made up by supplemental feeding or watering afterwards. Because supplemental feeding is not possible with a closed economy, the ratio values used in closed-economy experiments are always moderate, and to the left of the peak in the labor supply curve, on the rising part of the function. Over this range, the pattern in open and closed economies is the same (compare Figure 22 with Figure 26 at ratios of 50 and below). In addition to the range problem, it is also possible that the post-session supplemental feedings given in the open-economy situation may have a suppressive effect on behavior at high ratios, especially if the feedings are given soon after the end of the experimental session (this may be another effect of the proportional timing process discussed earlier).

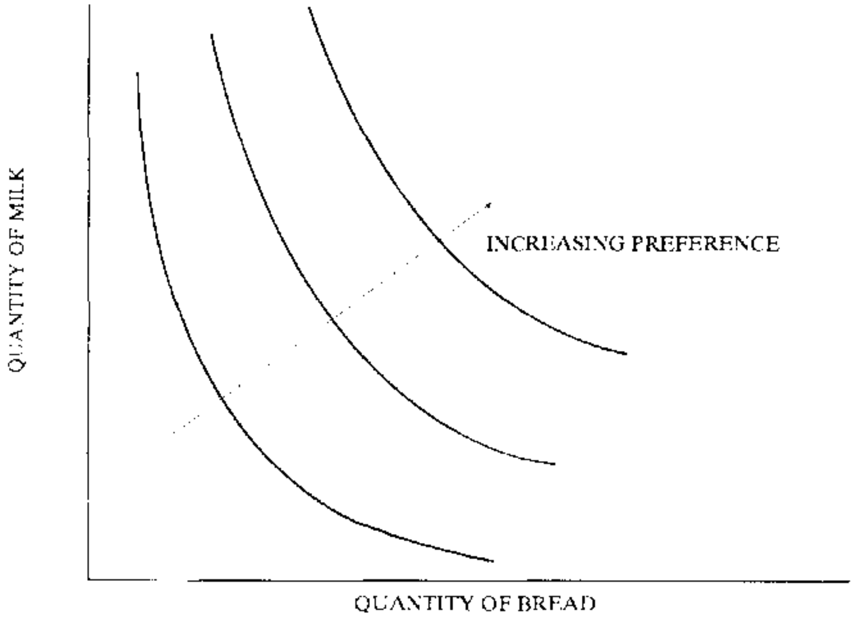
Notice that we cannot directly compare the absolute values of the ratios between open- and closed-economy experiments, because we cannot equate either the values of water and food reinforcement for the two species, or the relative effortfulness of pecking and lever pressing. The forms of the two curves, not the absolute values, are the features of interest.

If the "true" function relating response rate and ratio value is an inverted-U, does this mean that the labor-supply analogy is true and the demand law analogy false? The correct answer is that both are just *analogies*, and what we are interested in is *principles*: What common principles underlie both these functions, and do *they* apply to behavior on these schedules? In fact the common principles are just the same as the principles behind the optimality analyses we have already discussed, but refined in the particular context of economic analysis. I turn now to the concept of preference structure that economists have used to explain both the downward-sloping demand curve and the backward-bending labor-supply curve.

Preference Structure and Indifference-Curve Analysis

Economics is about value or, as economists term it, *utility*. Psychologists and philosophers differ on whether it makes sense to give numbers to utilities. Some psychologists say you can, most philosophers say you can't, and economists agree. Economists have devised a method that allows them to make predictions based only on value relations of "equal" and "greater than". The method works like this. Consider two goods, such as bread and milk. Even if we can't give numbers to the utility of a given quantity of milk or bread, everyone agrees that we can almost always *equate* the value of bundles of goods. For example, 2 quarts of milk and 3 loaves of bread may be judged equal to another bundle with 4 quarts of milk and 2 loaves, in the sense that we are indifferent as to which *commodity bundle* we get. There will be a whole set of bundles of this sort, differing in the proportions of bread and milk, but the same in that we are indifferent among them. This set defines an individual *indifference curve*.

Figure 27



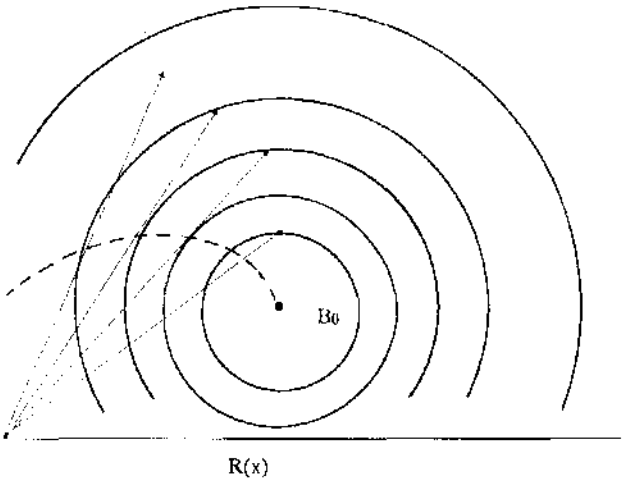
Conventional indifference curves.

An indifference curve represents a set of commodity bundles that are equal in value. Figure 27 shows several indifference curves. Any point on one curve is of equal value to all other points on the same curve. Everyone agrees that bundles of goods, or of activities, can be rank-ordered, A being preferred to B, B to C, and so on even if we cannot be sure by *how much* A is preferred. When bundles of goods are ranked in order of value, they occupy different indifference curves. For example, if we now compared 5 quarts of milk and 4 loaves of bread with 4 quarts and 2 loaves we would prefer the first. The 5,4 bundle must occupy a point on a *higher* indifference curve. Thus, a set of indifference curves is defined by the relation of value-equality among bundles on the same curve and greater than (or less than) for bundles on different curves.

The point of maximum value is represented by the point in the upper right corner of Figure 27. The *preference structure* for bread and milk, or any other pair of goods, can therefore be represented by a sort of contour map of indifference curves, such as those illustrated in Figure 27. The arrow indicates the direction of increasing preference which, for most real commodities (but not for activities) is generally in the direction of more of both goods. The analogy to contours is exact: like contours, indifference curves cannot cross one another. Like contours, they represent equal values on a third dimension: height, or value. A set of indifference curves is a contour map lacking a vertical scale: we know where the mountains are and which contour line is above which, but we don't know *how much* higher one is than another.

Knowledge of the preference structure is not by itself sufficient to predict behavior. We also need to know what constraints the organism is under. Once a constraint is specified, however, the optimal solution is clearly to settle on the highest indifference curve consistent with it. For reinforcement-schedule constraints, this implies an equilibrium at the point where the feedback function just touches is *tangent* to the highest indifference curve. This is illustrated in Figure 28 for choice between operant responding on the vertical axis and reinforcement on the horizontal axis. The indifference curves are circles centered on point B_0. Notice that B_0 is such that the rate of reinforcement is high, and of operant responding, low. The constraint lines are simply the feedback functions for different ratio schedules. The dashed line is the locus (path) of points traced out by a series of different ratio schedules in this space; this locus is termed the ratio-schedule *response function*.

Figure 28



Indifference curves derived from the minimum-distance model. Rays through the origin are ratio constraints. The dashed curve is the expected response location.

If you think about it a little, you will realize that circular indifference curves correspond to the minimum-distance model shown earlier in Figure 20 because circular curves correspond to the assumption that points equidistant from the free-behavior (bliss) point are all of equal cost. Since equal cost obviously corresponds to equal utility, it really makes no difference whether we do our analysis in terms of cost or utility; the indifference curves, and therefore the predictions, will be the same.

Notice that the response-function traced out in Figure 12 is an inverted-U shape. So also is the comparable function relating response rate to ratio value, which is not shown but you can see that response rate declines at high ratio values (steep feedback lines), rather than continuing to increase. In other words, the minimum-distance model predicts a labor-supply-type relation between response-rate (labor supply) and ratio value (wage rate). How does the response-cost model compare? We can easily derive indifference curves from the cost function (Equation 6),

$$C(x) = I + Qx,$$

where I is the interfood interval, Q is a cost parameter, and x is the rate of the operant response. Interfood interval is just 1/reinforcement rate, i.e., 1/R(x), so that the equation becomes

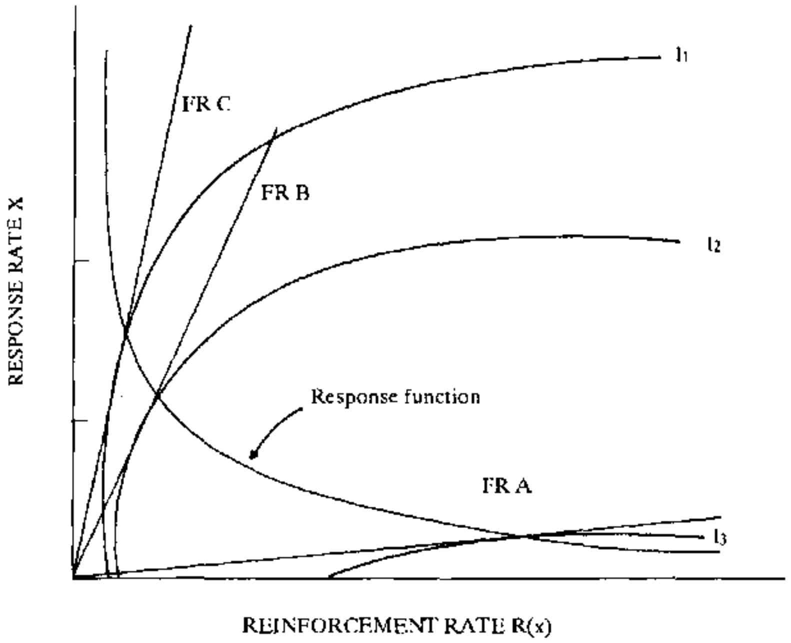
$$C(x)= 1/R(x)+Qx. \tag{7}$$

To plot an indifference curve, we just pick a value for the cost, say C, set the right hand side of Equation 7 equal to it, and rearrange so as to get x as a function of R(x). The result is

$$x= [C-1/R(x)]/Q, \tag{8}$$

which we can then go ahead and plot. Each time we change the value of C, we get another indifference curve. Indifference curves for three values of C are plotted in Figure 29. The response function goes through the points at which the three ratio-schedule feedback functions are tangent to the three indifference curves. As I showed earlier, the RC response function slopes downwards to the right, like a demand curve.

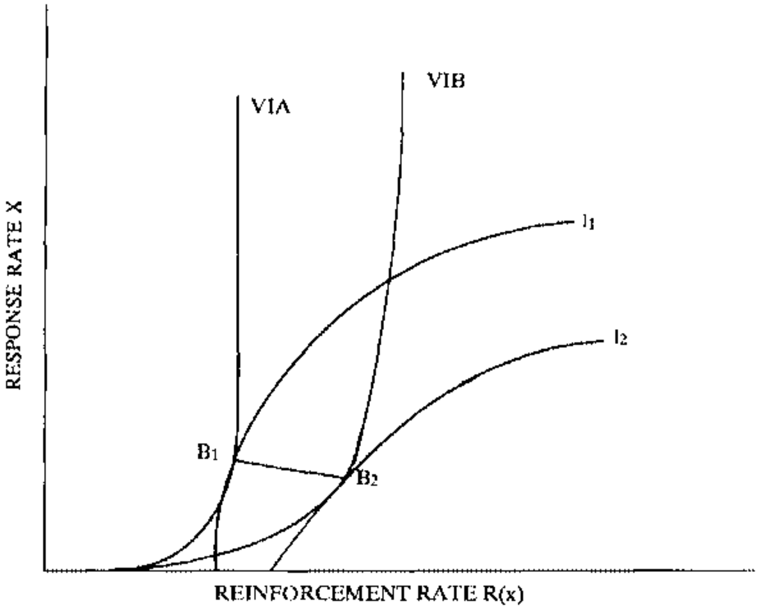
Figure 29



Indifference curves for the response-cost model (equation 8 in the text with C = 3, 10 and 15; Q = 0.5). Feedback functions for three ratio values are shown. The downward-sloping response function is indicated by the two line segments.

The nice thing about indifference curve analysis is that it makes it easy to separate the cost function, which is the essence of the model, from the constraints. The set of indifference curves represents the cost function; the constraint lines then define on which indifference curve, and where on the curve, behavior will lie. I have only described ratio-schedule constraints so far, but it is almost as easy to derive predictions for other schedules, such as variable-interval. Figure 30 shows two indifference curves for the RC model that are tangent to two VI feedback functions. Both points of tangency are at the same response rate, in agreement with the optimal VI policy I derived earlier from the RC model.

Figure 30

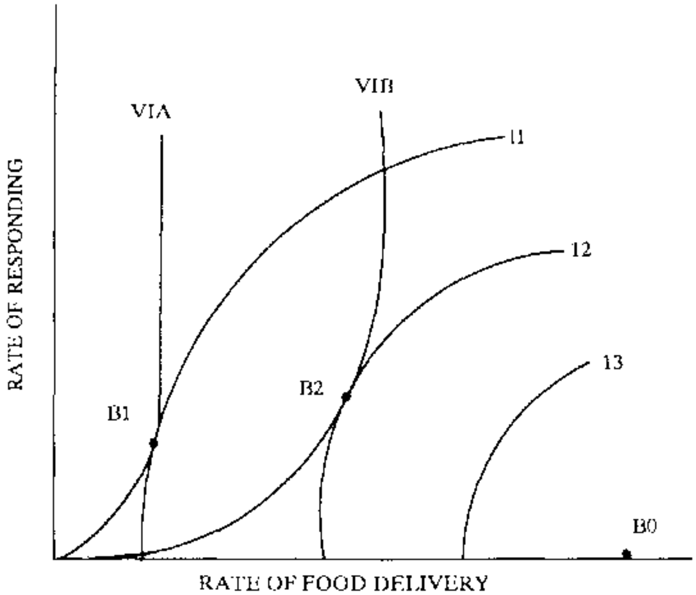


Indifference curves for the response-cost model (curves I_1 and I_2) and two VI molar feedback functions (VI A and VI B) tangent to them. B_1 and B_2 are two points on the response function.

Notice that the VI feedback functions in Figure 30 have a shape that makes sense, based on what you know of VI schedules: when response rate is very low, reinforcement rate is almost proportional to response rate. On a VI 1-min schedule, for example, if an animal ups his response rate from once per hour to twice per hour, his reinforcement rate will approximately double. But as response rate approaches the maximum reinforcement rate specified by the schedule, reinforcement rate rises more and more slowly, approaching as an *asymptote* the maximum rate permitted by the VI.

Now we are in a position to see the real difference between the labor-supply and demand curve analogies. The real difference is in the form of preference structure they each imply. The inverted-U labor-supply function implies indifference curves resembling those of the MD model; the downward sloping demand curve function implies indifference curves like those of the response-cost model. As you can see in Figure 31, the indifference curves for the MD model differ slightly from those of the RC model in such a way as to produce a VI-schedule response function that rises over most of the range (response rate at point B2>B1), rather than being constant, as in Figure 30. On balance, the MD analysis and related economic models of Rachlin and his associates (e.g., Rachlin et al., 1981) fit experimental results a bit better than the response-cost model.

Figure 31



Indifference curves for the minimum-distance model and VI molar feedback functions tangent to two of them. B1 and B2 are two points on the response function.

All the optimality analyses we have discussed so far interfood interval minimization, response cost, and minimum-distance imply a certain preference structure. As you have just seen, given the form of the cost function, we can easily derive the form of indifference curves. So why do we need yet another method for doing the same thing? What can we do with an indifference-curve analysis that we can't do using a cost function?

First, the indifference curve method is potentially *experimental*: it tells us (at least in principle) how to actually measure the organism's preference structure by equating commodity bundles. This is known as the method of *revealed preference* (Samuelson, 1965), which is the closest that economists come to being pure behaviorists. Like reinforcement theorists, they are almost unanimous in agreeing that the only way to assess value is through the individual's expressed preferences, represented by indifference curves. In fact, neither economists nor psychologists much use the direct method of asking people or animals to equate commodity bundles. Although there is now a small school of experimental economists, the majority do no experiments whatever. And psychologists, for practical reasons, usually do experiments of the type we have already discussed, and then test economic models indirectly.

Second, a set of indifference curves need not follow any particular mathematical form. It can be anything we like, subject to the logical limitation that indifference curves, like contour lines, cannot cross. While some behavior theorists (e.g., Battalio, Kagel, Lea, Staddon, most behavioral ecologists) have favored preference structures derived formally, from assumptions such as minimum distance or response cost, others (e.g., Hursh, Rachlin, many economists) have favored the greater flexibility of a graphical approach in which indifference curves are simply drawn in a plausible fashion.

Third, because indifference curves represent the cost function directly, they show the real similarities between models. You can see this easily by comparing Figures 30 and 31. The indifference curves derived from minimum-distance and response-cost, are clearly very similar, even though the equations that describe the two models do not appear similar at all. Hence, the models are in fact quite similar, despite the very different defining assumptions.

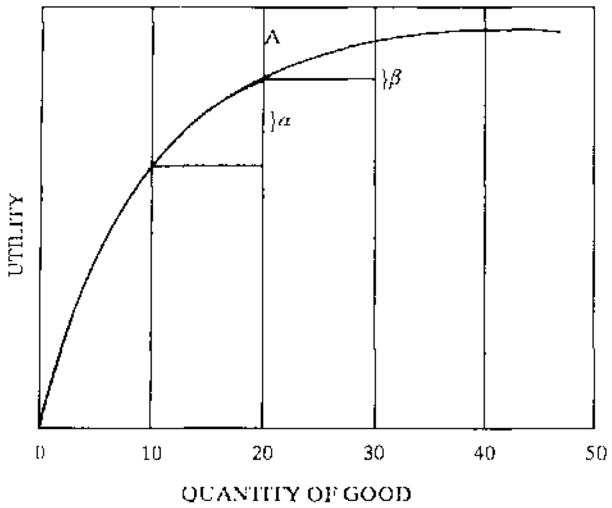
I turn now to a discussion of what different forms of indifference curve imply for behavioral allocation and choice.

Marginal Value and Substitutability

It is no accident that the indifference curves we have been discussing are all *convex*, that is, they curve outwards from the region of maximum value.

Convexity represents a very important property of cost, usually expressed as the *law of diminishing marginal utility* (or increasing marginal cost). What it means is that the more you have of something, the less each additional increment is worth (Conversely, the more you lose of something, the greater the cost of each additional bit of loss).

Figure 32



A negatively accelerated utility function. α and ß are successive increments in utility produced by equal increments in the quantity of a good.

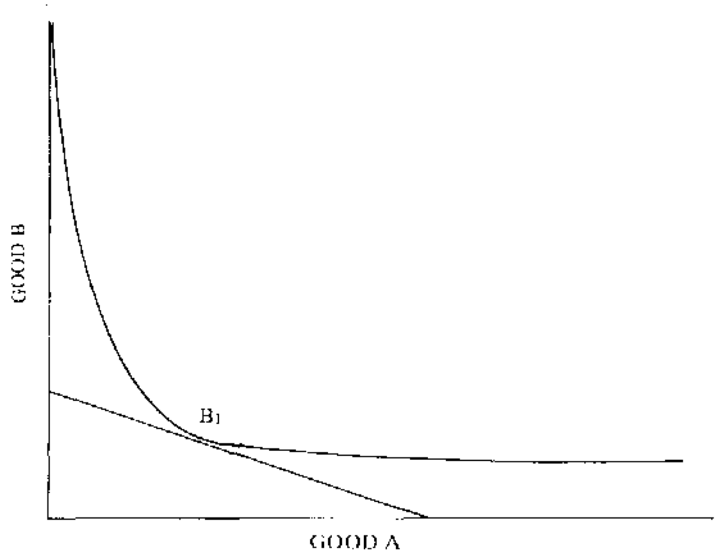
Diminishing marginal utility is illustrated in Figure 32, which shows a negatively accelerated utility function. The vertical lines indicate increments of the good (e.g., food-rate increments). The two little vertical lines show you two increments in utility associated with successive equal increments in the good. As you can see, the first utility increment, α, is greater than the next, ß, and this pattern continues across the whole curve. Thus, going up and to the right from point A (increasing utility), the curve shows diminishing marginal utility; going down and to the left from A, the same curve shows *increasing* marginal cost. Diminishing marginal utility is sometimes termed *satiation* and there is a close relation between this property and the dynamic property of satiation that we discussed earlier: both imply behavioral diversity, as we will see in a moment.

The concept of marginal utility is important because it tells us how to allocate resources for maximum satisfaction. For example, given $100 to spend on any mixture of three commodities, bread, milk and beer, maximimum satisfaction is assured if we are indifferent on whether to spend our last dime on more milk, bread or beer. Indifference means that we have equated the marginal values of the three commodities. Those familiar with differential calculus will recognize that equating marginal value is just the mathematical operation of equating the partial derivatives of the value (or cost) function with respect to each commodity.

Marginal value can be understood without getting into higher mathematics. For example, consider how you should allocate a fixed total amount of time between two activities whose utilities are each proportional to time spent not curvilinear, like the utility function in Figure 32. Clearly, the best strategy is to devote all the time to one activity or the other, depending on which line has the steeper slope. It is more typical, however, for value functions to be *negatively accelerated*, as in Figure 32, in which case, the more one does of something, the smaller the marginal benefit. For example, the more you eat of something the less tasty each additional morsel becomes. It is pretty obvious that most goods show diminishing marginal utility: the first ice cream is wonderful, the next is nice the tenth may make you sick. Moreover, if two goods show diminishing marginal utility, then when you must allocate limited resources among them, your optimal policy is spend some money on both, i.e., to show a *partial preference*, rather than exclusive choice of one or the other. If the goods are activities, each of which shows diminishing marginal utility, then clearly the optimal policy is to spend some time in each activity, to show behavioral diversity, rather than behavioral stereotypy.

The relation between partial preference, diminishing marginal utility, and convex indifference curves, is illustrated in Figure 33. The figure shows an indifference curve for two goods, A and B, each of which shows diminishing marginal utility of the form shown in the previous figure. As you can see, the curve is convex. The diagonal line represents what economists call the *budget constraint*: a fixed sum of money, which must be spent on A and B. The slope of the line represents the relative prices of A and B: if both prices are equal, the line will intersect both axes at the same value. If A and B are activities, measured as proportions of *time* spent, the budget constraint corresponds to the time-allocation constraint we discussed earlier. The optimal allocation of resources is at point B_1, where the constraint line is tangent to the indifference curve. B_1 obviously represents a partial preference.

Figure 33



An indifference curve for two partially substitutable goods.

What will be the form of indifference curve when utilities do *not* show diminishing returns? In the example I just gave, the utility of each activity was simply proportional to the amount of activity: $U_1=aT_1$ and $U_2=bT_2$, where U_1 and U_2 are the utilities of investments, T_1 and T_2 are the times invested, and a and b are the ratios of utility to time spent. Now recall how we derive indifference curves from utility (or cost) functions. We pick a total utility (or cost), call it C, and set the combined utilities equal to it, so $C=U_1+U_2=aT_1+bT_2$; rearranging gives us $T_1=(C-bt_2)/a$, which is a straight line. There is obviously no way that a straight-line budget constraint can be tangent to a straight-line indifference curve. Hence, the optimal policy is to choose activity 1 or activity 2 exclusively, as i already showed.

Note that a straight-line indifference curve need not imply that the goods it relates are not subject to diminishing marginal utility. It may also mean that the two goods are not independent, so that consumption of one causes satiation for both.

Economists have names for these different kinds of indifference curves: Two goods related by a straight-line indifference curve, where the optimal policy is exclusive choice, are termed *perfect substitutes*. Examples are two equivalent brands of gasoline: you just pick the one with the lowest price.

Two goods related by a convex indifference curve are termed *partial substitutes*, because a change in price (slope of the budget line) causes a change in preference that falls short of a complete switch from one to the other. Coffee and tea are partial substitutes: if tea becomes more expensive, you may drink more coffee. Two goods related by an extreme convex indifference curve (i.e., a corner, made up of a vertical and a horizontal line) are termed *complements*. The idea here is that there is a preferred ratio of the two goods (e.g., equal numbers of left and right shoes), so that we are indifferent to additional quantities of either good: a bundle of 105 left shoes and 100 right shows has the same value as a 100:100 bundle.

These names are not very helpful for analysing the kind of behavior under discussion, because the terms say more than they mean. *Substitution*, for example, implies *functional equivalence* (i.e., that consumption of one good satiates for both), but all it means is a particular form for the indifference curve. As we have seen, a given indifference curve may be generated in several ways, only some of which correspond to functional equivalence. But since these terms are in common use, you need to know something about them.

Implications for Behavioral Allocation and Choice

I have already pointed out the most obvious implication of imperfect substitutability for implicit choice: partial substitutability implies behavioral diversity. These ideas can also help us understand the results of explicit choice experiments in which the reinforcers are of different types or delivered in different amounts. Let's consider first one of the simplest choice situations, the two-armed bandit, and the effect of different types of reward.

A *two-armed bandit* is just a two-choice version of the familiar Las Vegas *one-armed bandit*, that is, a situation in which the animal has two choices (levers, response keys), each of which delivers reward on a probabilistic (i.e., random-ratio) schedule. For example, our subject may be a rat responding on one lever for Cherry Cola, and on another either for the same thing (Cherry Cola) or for something different (Tom Collins mix). Suppose the two random-ratio schedules are the same, e.g., 25, so that the rat gets access to either reinforcer with probability 0.04 for pressing the appropriate lever. What should he do if both levers produce Cherry Cola? Well, it really doesn't matter, so that the chances are that after much experience, our rat will just develop a *position preference* and respond exclusively on one or other lever. This outcome is practically guaranteed if the two ratio schedules are *un*equal, e.g., 15 and 25: most rats will eventually fixate on the higher-probability lever.

This is just what we would expect with two rewards that are perfect substitutes. If the animal has a fixed number of responses to expend, then the constraint line is just like the budget line in Figure 17: x responses on the Left means N-x on the Right, where N is the total permitted, and the prices (ratio values) are equal. If the two rewards are perfect substitutes, then the indifference curve is also a straight line, so that the prediction will almost always be exclusive choice of one option or the other.

What will happen if the rewards for the two responses are different (Cherry Cola and Tom Collins mix)? Now the indifference curve may be convex-complete satiation on Cherry Cola may leave our rat still with some appetite for Tom Collins mix, and vice versa. The prediction for equal ratio schedules is now very different: partial preference, rather than exclusive choice. Moreover, if the partial preference favors Cherry Cola, say, then we can increase the ratio on that side without abolishing responding-indeed, the change in preference for a given change in "cost" allows us to estimate the indifference curve directly. It turns out that the behavior of rats is consistent with this simple economic analysis (Rachlin et al,. 1981).

Now let's look at different food-reinforcer *amounts* for each choice, with equal ratio values for each. No matter what the form of the utility curve for food amount, so long as more is preferable to less, our rational rat should always pick the large-amount option exclusively, and rats (after sufficient experience) usually do. But we can see the effect of the utility function if we change the procedure slightly: on the Left, the animal continues to get a small amount, say a 20 mg food pellet. But on the Right we give him either a very small (10 mg), or a large (30 mg) pellet, with equal probability. So now 25 responses on the Left on average buys our rat 20 mg of food; 25 responses on the Right buys him either 10 or 30 mg, which averages out to the same amount. The question is, which should he prefer? The answer depends on how the utility of food depends on its amount.

Look again at the negatively accelerated utility curve in Figure 32. Let's see how much utility (as opposed to how much food) our rat gets with the two options we have described. On the Left, the rat gets amount A in the Figure, about 0.8 on the utility axis. On the Right he gets either $0.8-\alpha$, or $0.8+\beta$. But since α is plainly always greater than ß (given a negatively accelerated utility function), then the average of $0.8-\alpha$ and 0.8+ß must be less than 0.8 so the rat should prefer the side with the fixed food amount to the side with variable amounts with the same average. And rats do. This aversion to variability is known to decision theorists as *risk aversion*, and it is a very common result whether the decision agents are people or animals. This explanation for risk aversion is a very old one, as psychological explanations go, being first offered by the Swiss mathematician Daniel Bernoulli in 1738.

Risk aversion can easily be demonstrated with human subjects. For example, the two psychologists Daniel Kahneman and Amos Tversky (1979) have become famous for a series of experiments in which they asked college students deceptively simple questions about decisions involving gambles and got surprising answers. In one experiment, students were asked to decide between the following two outcomes:

Which of the following would you prefer: $3000 for sure, or a 0.8 chance of $4000?

Since 0.8 x 4000=3200, the gamble has the higher expected value. Nevertheless, the great majority of subjects opted for the sure thing.

Risk aversion is evidently a widespread characteristic. It makes good adaptive sense, for at least two reasons. First, the future is uncertain: "a bird in the hand is worth two in the bush". Second, the value of reinforcements such as food depends not just on their amount but also on their distribution in time. An individual may consume 100 kg of food per year, say, but he is unlikely to trade a regimen of 0.3 kg/day (total: 109.5 kg), for 200 kg delivered at year's end. A large food reward should be less valuable than two half-size rewards delivered twice as often, although the difference might well be small if the amounts and delays are also small. Since any repeated gamble involves a change in temporal distribution, as well as distribution of amounts, risk aversion will often be adaptive.

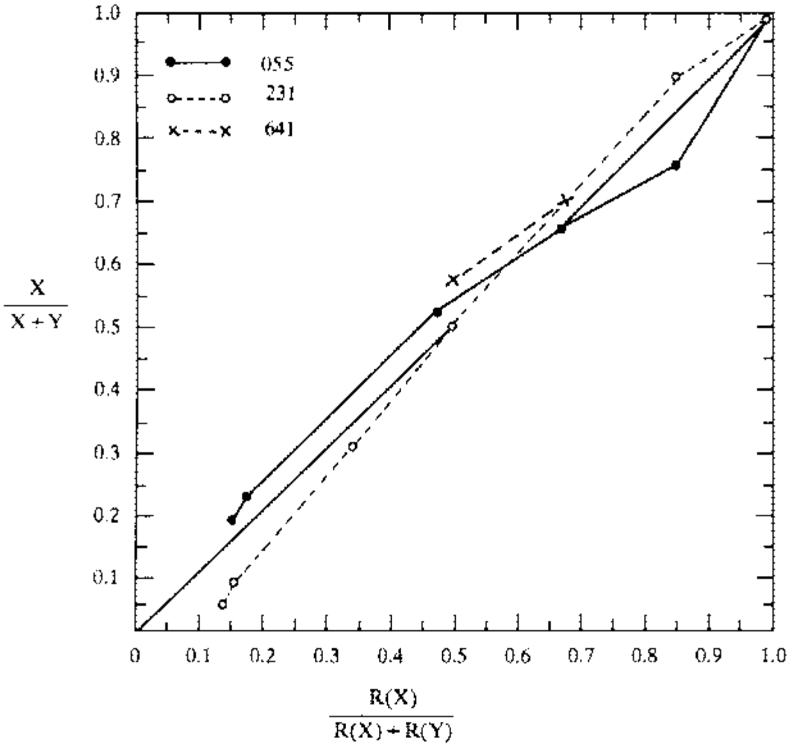
The Matching Law. Convex indifference curves, and linear constraints, yield partial preference. Let's look again at variable-interval schedules and see how this rule might apply when an animal has to choose between two independent VI schedules. Figure 30 shows how the reinforcement rate obtained on VI depends on the response rate. If you can imagine the axes reversed, with response rate on the x-axis and reinforcement rate on the y-axis, you can see that the VI feedback function has exactly the same form as the negatively accelerated utility function in Figure 32. As responding increases from zero, reinforcement rate at first rises rapidly, but then more slowly as it approaches the maximum prescribed by the VI schedule. Let's see if we can figure out what an animal should do when confronted with two concurrent variable-interval schedules. We'll assume the simplest possible utility function: average reinforcement rate. The animal is to allocate his responding to get as many reinforcements as possible within a fixed time. Given a fixed total number of responses to "spend", the animal should spend them so that the marginal reinforcement-rate gain on each of the two schedules is equal. Since the two feedback functions are both negatively accelerated, this must yield a partial preference: some responding on both VI schedules, more on the richer one. In fact, with the molar VI feedback function we have been using, reinforcement-rate maximization of this sort predicts that the animal should *match* the ratio of responses made to reinforcements obtained: $x/y = R(x)/R(y)$, where x and y are the total number of responses to Left and Right, and $R(x)$ and $R(y)$ are the total number of reinforcements obtained (Staddon & Motheral, 1978). In other words, if we find that our rational pigeon got 40 reinforcements for pecking on the Right and 20 for pecking on the Left, we expect that he will have made twice as many pecks on the Right as on the Left. Rewritten in terms of proportions, this is the well-known *matching law* for responding on concurrent VI VI schedules

$$x/(x+y) = R(x)/[R(x) + R(y)]. \tag{9}$$

The matching law was discovered by Richard Herrnstein (1961), and his original data, taken from pigeons pecking two keys for food reinforcement delivered on VI schedules, are shown in Figure 34. Each data point plots the proportion of Left and Right key pecks against the proportion of reinforcements obtained after several weeks of daily exposure to a given pair of VI values. To minimize changes in overall response rate, the total rate of food delivery for Left and Right pecks combined was held approximately constant. The diagonal line indicates perfect matching, and as you can see the data from the three pigeons conform closely to it.

The matching law has been the focus for an extraordinary amount of experimental and theoretical work in the more than 25 years since it was first discovered. The law itself, originally just an empirical finding, has been used as the basis for a molar theory for reinforcement-schedule performance in general, as we will see in a moment.

Figure 34



Results from three pigeons on concurrent VI VI schedules. Each point represents a different condition. The diagonal line represents perfect matching (from Herrnstein, 1961).

The derivation of matching from simple reinforcement-rate maximization gave a great boost to the optimality approach. It turns out that a great many maximization models predict matching, because of the negatively accelerated form of the VI molar feedback function. Closer study soon revealed difficulties, however. As we have seen in all our earlier discussions, optimality models always fail under some conditions because they are functional not mechanistic, models. Molar reinforcement-rate maximization fails to predict matching in a situation closely related to concurrent VI VI: concurrent VI VR. When one alternative is a VI but the other is a variable-ratio, pigeons spend too much time on the ratio alternative: maximization predicts more responding on the ratio (where responses count more) than on the VI (where they count less). Nevertheless, pigeons, blindly obedient (apparently) to the matching principle, continue to match response ratios to ratios of obtained reinforcements, though with a slight *bias* in favor of the ratio: $x/y = kR(x)/R(y)$, where k is greater than one, and X is the ratio-reinforced response (Herrnstein & Heyman, 1979).

Stephen Lea (1978) has investigated matching and maximizing in a more subtle way. His pigeons could respond on either of two keys: on the Left (say) they got food according to a random-ratio schedule with constant probability, p. On the Right, they got food according to an *adjusting* ratio: with each food delivery on the Right, the ratio increased; with each food delivery on the Left, the ratio on the Right decreased. This is termed a *titration* procedure (we have already encountered a titration schedule in an experiment by Mazur, discussed earlier). The point is that the optimal strategy here is for the animal to work for some time on the Left, until the ratio on the Right has reached a low value; then to switch and collect a few "cheap" reinforcements before switching back to the Right and repeating the process. Instead of maximizing in this way, Lea's pigeons alternated back and forth in such a way that the net payoff probability on the Right (adjusting) side matched the probability on the Left (constant). If we denote payoff probability on the Right by q and on the Left by p, then $p = q$ represents the birds' steady-state behavior. But p is just equal to the ratio of reinforcements to responses on the Left, $R(x)/x$, and similarly for q, so that Lea's result is just another example of the matching law: $x/R(x) = y/R(y)$, or $x/y = R(x)/R(y)$.

Although Lea's data disprove molar maximizing as an account of choice, they are perfectly consistent with what has been termed *local* or *molecular maximizing*. If the pigeon just allocates his pecking on a moment-by-moment basis to the alternative that offers the highest probability of reinforcement, then the titration schedule ensures that the average payoff probability will be the same on both sides, which implies matching. This very same molecular maximizing process was suggested many years ago as the basis for matching on concurrent VI VI schedules (Shimp, 1966). We will return to molecular processes in a moment. But first, let's see if there is any good evidence that marginal changes in molar variables such as reinforcement rate have direct effects on behavior.

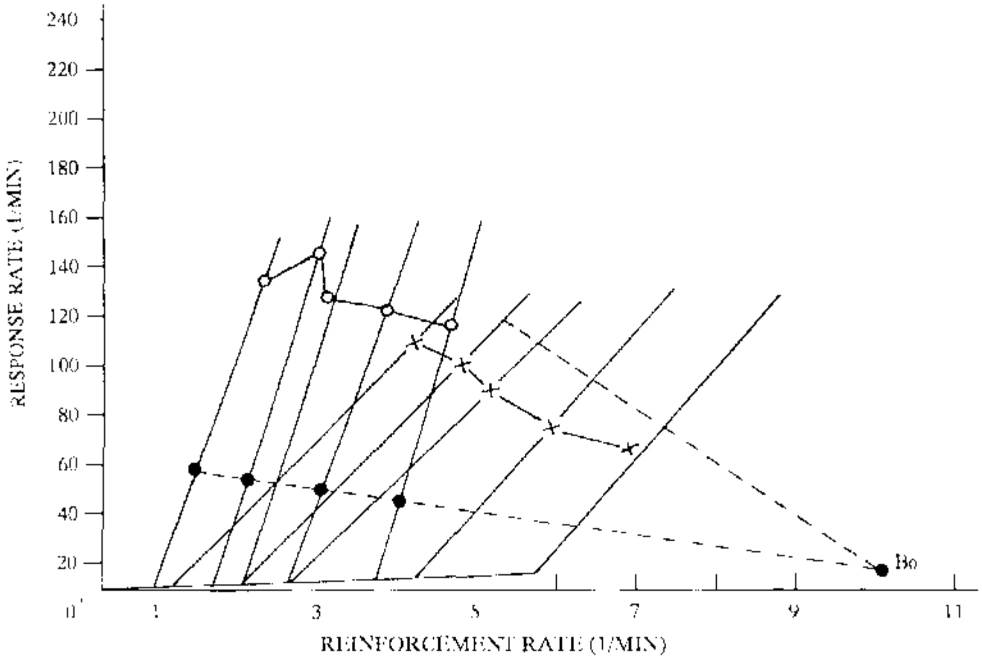
Is Behavior Directly Sensitive to Marginal Molar Changes?

The fundamental assumption of molar optimality is that behavior is sensitive to marginal quantities, measured at the molar level. How valid is this assumption? Ettinger, Reid and Staddon (1987) recently carried out a direct test. They chose a schedule that has linear molar feedback functions, because many optimality models predict a particularly simple adaptation to such schedules: a straight-line response function. These models all predict that a change in the *slope* of the molar feedback function (i.e., a change in its marginal rate of return) should always cause some change in the slope of the response function.

Ettinger at al. used what are called *interlocking* schedules. An interlocking schedule is a combination interval-ratio schedule. If the animal does nothing, the schedule is effectively a fixed-interval. But the inter food time is reduced by every response, as it is on a ratio schedule. For example, if the animal does nothing, food becomes available for the first response after 60 s, say. But if the animal makes one response before the 60 s mark, then food becomes available after 55 s; if he makes two, food is available after 50 s, and so on. Thus, the interfood interval is determined according to the formula $I = T-am$, where m is the number of responses that are actually made, (excluding the response that actually procures the reinforcer), T is the minimum interfood interval if no responses are made, and a is a parameter that says how much each response reduces the time to food.

The results of one of the experiments by Ettinger et al. are shown graphically in Figure 35. The figure shows feedback functions for two sets of interlocking schedules that differed in the slope parameter, T (two sets of positive-slope lines). Within each group, the slopes were the same but the intercepts (parameter a) differed. Response functions predicted from the MD model are sketched in as dashed lines. Other optimality models predict different changes in the slope of response function. The point is that all molar optimality models predict *some* change in slope between the two conditions. Yet the results are clear: (a) the obtained response functions are indeed approximately linear, but (b) there seems in fact to be only one linear function, which is the *same* for both sets of schedules. All the data points seem to lie on the same straight line with negative slope. The difference in molar marginal rate of reinforcement between the two series evidently had no effect on response-function slope. These data provide a very strong hint that here, as with the chain schedules discussed earlier and Lea's titration schedule, we need to look at the local, molecular level to understand how reinforcement schedules actually affect behavior.

Figure 35



Results from one animal on two sets of interlocking schedules. The interlocking schedules are represented by the two sets of five lines with positive slopes. The minimum-distance predictions are represented by dashed lines. The obtained response functions are represented by the solid lines that connect the symbols (from Ettinger, Reid, & Staddon, 1987).

Molar Optimality: Conclusion

What can we conclude about the status of molar optimality as a valid model for implicit and explicit choice? We must first acknowledge the general conclusion that no optimality model works in every situation. Animals are rarely, if ever, "literal optimizers"; they don't remember the average payoff associated with a given pattern of responding and compare it with the payoffs for other patterns (from a well-defined set of possibilities) and then pick the best pattern, as some views of optimal responding seem often to imply. Under most conditions, people don't behave in this way either. Moreover, the data from the Ettinger et al. experiment strongly suggest that the real causal relations are not at the molar level at all. Nevertheless, the simple local rules animals often use perform remarkably well in a wide range of situations. Hence, molar optimality analyses also do rather well. Optimality analyses have also revealed unsuspected relationships between implicit and explicit choice, and they still provide the simplest account for choice between partially substitutable reinforcers and for the analysis of choice behavior under risk. Optimality models show unequivocally that reward value is a negatively accelerated function of reward amount, so that doubling amount increases value by a factor less than two. Recall that we needed to assume a less-than-proportional relation between reward size and reward effect in order to get correct predictions of self-control results from the proportional timing model. In short, optimality models provide a tolerably good picture of what animals manage to achieve, but a poor picture of how they actually do it. Animals often behave optimally they almost never optimize in any literal sense (Staddon & Hinson, 1983).

Behavioral economics is just optimality theory by another name, and with a slightly different bag of theoretical tricks: indifference curves instead of objective and cost functions. But the rationale, the results, and the limitations of behavioral economics are exactly the same as those of optimality. Only the names and connotations are different. "Labor supply" and "demand curve" suggest different things than "response cost" and "minimum distance" but the principles behind them are the same.

We turn now to a different kind of molar approach, which we term *molar descriptive theories* of behavioral allocation and choice. The final section of the chapter deals with molecular and dynamic analyses.

Generalized Matching: The Relative Law of Effect

The simplicity and reliability of the matching law suggested to Richard Herrnstein and his students that it might well form the basis for a general theory of behavior on reinforcement schedules. Recall the form of law, stated in terms of proportions (Equation 9): $x/(x+y) = R(x)/[R(x)+R(y)]$. Suppose we replace the denominator on the left hand side with a term representing all behavior in a given situation, and the denominator on the right with a term representing *all* sources of reinforcement:

$$x_1/Ox_i = R(x_1)/OR(x_i),$$

where x_1 is one response, Ox_i is the sum of all responses and $OR(x_i)$ is the sum of all reinforcements. If we add the single assumption that the sum of all responses is constant, then bring the constant over to the other side, the equation reduces to

$$x_1 = kR(x_1)/OR(x_i), \tag{10}$$

where k is now a constant of proportionality. In words, it says that the rate of a given response, x_1, is proportional to its rate of reinforcement *relative* to the reinforcement for all other responses. Herrnstein (1970) called this the *relative law of effect*.

Herrnstein had to make two assumptions to get from the simple matching result to the relative law. The first is to assume the constancy of the term Oxi, the sum of all responses. The second is to assume that all sources of competition are also sources of reinforcement. Activities such as grooming or sleeping, which take away time from the operant response, must therefore be treated as sources of reinforcement. The latter assumption is hard to verify, and probably unnecessary to the theory as a formal description of results. The first assumption is wrong unless each activity is scaled in terms of the time it takes up. The constant, k, in Equation 10 therefore does double duty: it represents the total period of observation, divided by the duration of one instance of activity X_1.

The simplicity of the relative law of effect has proven very attractive, and strenuous attempts have been made to extend it to a wide variety of schedules. Here we will just consider its extension to responding on single-response (i.e., not concurrent) VI schedules. In this case the term $OR(x_i)$ simplifies to $R(x) + R_0$, where R_0 is just the sum of reinforcers for activities other than response X, so that Equation 10 for a single response is

$$x = kR(x)/[R(x) + R_0], \tag{11}$$

where x is response rate and R(x) reinforcement rate, as usual.

Let's look a little more closely at Equation 11, which sometimes known as *Herrnstein's hyperbola*. When R(x) is low relative to R_0 (i.e., a very low reinforcement rate for X), the equation reduces to $x = kR(x)/R_0$, i.e., response rate is proportional to reinforcement rate. Conversely, when R(x) is large relative to R_0, the equation reduces to x = k. These two properties should by now be very familiar. They correspond to the negatively accelerated pattern of response rate vs. reinforcement rate we have already seen for pigeons and rats on VI schedules. You have already seen examples in Figures 18 and 19. It seems that the quantitative law of effect provides an accurate account of response rate vs. reinforcement rate functions not only for concurrent VI VI schedules, but also for simple VI. How far can this idea be extended?

Herrnstein has proposed extensions to explain behavioral contrast on multiple schedules, but these extensions have not been well supported by experimental work. The principle also fails on simple ratio schedules and interlocking schedules. As we saw earlier, response rate on ratio schedules is either a declining function of reinforcement rate, or an inverted-U; response functions on interlocking schedules follow the linear declining form shown in Figure 35. The problem is that Equation 11 has no provision for the reinforcement schedule like Allison's conservation theory, it predicts the same thing for any reinforcement schedule. So the relative law of effect has failed to provide us with a general principle of reinforcement. Nevertheless, the function it predicts for simple VI schedules (Equation 11) is accurate. In a moment I will suggest a basis for Eq. 11 that is different from Herrnstein's.

Molecular mechanisms of behavioral allocation

Matching of response ratios to reinforcement ratios is a very robust finding on concurrent VI VI and concurrent VI VR schedules. That's why the matching law was proposed as a general reinforcement principle. Yet the attempt failed. Can there be other reasons why matching is such a reliable result? We have two places to look: in the animal, and in the procedure. Perhaps the animal is following some simple local rule that results in matching? Perhaps the procedure is such that many patterns of behavior inevitably result in matching? Let's look at the procedure first.

One of the basic properties of behavior is *variation*. Variation serves many adaptive functions, but one of the most important is that it helps animals to detect regularities. Pigeons in choice experiments show some moment-by-moment variation in both their overall rate of responding and their allocation of responses to each key. What effects might this variation have, given the known properties of VI schedules? To answer this question, we need to go back to the feedback function for VI schedules. The VI feedback function has two general properties: when response rate is very low, the schedule is in effect a fixed-ratio 1 schedule (i.e., every response is reinforced); and when response rate is high, reinforcement rate is almost constant. Consequently, when our pigeon responds very slowly on each key, he will get food for almost every response. If he gets food for every response on each key, then he *must* match response ratios to (obtained-) reinforcement ratios. We can safely conclude, therefore, that when response rate is low, matching is an *artifact* of the properties of the VI schedule.

This possibility has been well known for many years. Proponents of matching as a general principle have therefore focused on the second possibility offered by VI schedules, namely that when response rate is high, reinforcement rate is almost constant and therefore independent of response rate. But the reality of behavioral variation means that we cannot assume a constant, high response rate. We must assume a range of rates, as well as some bias in favor of the higher reinforcement-rate alternative. The issue is quantitative and involves at least two properties of behavior: its amount (how fast does the pigeon peck?), and its bias (how much does he prefer the better key?). It is not something easily settled by verbal arguments. But we can get some insights into what is going on with the aid of *simulation*: writing a computer program to "respond" at various rates with different biases and deliver "reinforcers" at random times. Simulations (e.g., Hinson & Staddon, 1983b) have shown that on concurrent VI VI schedules, at least, almost any pattern of responding that satisfies two conditions will result in matching. The two conditions are (a) that response rate on a given side is some positive function of the rate of reinforcement obtained, and (b) that if no reinforcement is obtained, response rate is zero. Both these properties of operant behavior were well established long before the matching result was obtained. Hence, the matching law has been valuable more for the research it has stimulated than for what it is has revealed about the actual mechanisms of reinforcement. It is a law about the behavior of a system animal plus schedule rather than about the animal itself.

If animals are not molar matchers, then what are they? In this final section we discuss the molecular mechanisms of choice that may underlie all the molar patterns we have been discussing. The first two, momentary maximizing and melioration, are examples of hill-climbing processes, that is, processes that move behavior in the direction of the better alternative. The last, linear waiting, is a slight extension of the proportional timing process discussed earlier.

Momentary Maximizing

We have already seen that animals will usually settle for the high-probability alternative in the two-armed-bandit situation. Given one key that pays off with probability 1/10 and another that pays off with probability 1/20, pigeons will eventually learn to peck only the high-probability key (e.g., Herrnstein & Loveland, 1975). Charles Shimp in 1966 proposed that pigeons follow this rule consistently, even on VI schedules where it does not seem to apply in any obvious way. He called his hypothesis *momentary maximizing* (MM). How does momentary maximizing work? The simplest way to understand MM is to begin with a procedure very similar to the concurrent VI VI schedule: concurrent FI FI. I discussed concurrent FI FI earlier (see Figure 6).

Recall that the concurrent FI FI could be programmed in two ways, either the two programming timers both reset with reinforcement for either response, or they are independent, the timer on the Left only resetting following reinforcement for a Left response, and similarly for the Right. Recall also that the latter method, independent timers, favors responding to both keys partial preference whereas the other method favors exclusive choice. Now it's time to consider why. Let's take a simple example, a concurrent FI 3-min FI 1-min schedule. Assume that our pigeon is initially naive and knows nothing about either alternative, other than being willing to peck both with no preference. This behavior gets him food after 1 min on the Right, which may cause him to shift his preference a little to the Right. So he gets a second reinforcement on the Right after 2 min. But presumably two food deliveries are insufficient to produce an exclusive preference, so he will still peck Left once in a while. After 3 min, he gets a third reinforcement on the Right, but now his next peck on the Left also procures food, because the FI 3-min timer has been running right along, and has just now "timed out". Now if we were to continue to watch the pigeon, we might begin to see more complex patterns develop. If he can keep both sides "in synch", then he can learn that a peck on the Left is only reinforced some time after two reinforcements on the Right, so we might see the development of a sequential pattern. But even without such a pattern, it should be obvious that our bird will continue to peck on both keys.

Now let's complicate the situation a bit further. Instead of concurrent FI FI, let's look at concurrent VI VI. The general argument still applies. The pigeon should clearly sample both alternatives, but now there is absolutely no possibility of learning any kind of response sequence like "peck on the Right until you get two reinforcements, then try the Left", because reinforcers occur at unpredictable times on both sides, so that two Left reinforcers may sometimes occur in close succession. What information might the pigeon use to help him?

Shimp's insight was to see that although reinforcement is always uncertain on a VI schedule, the *probability* of reinforcement changes in a predictable way. As Staddon, Hinson and Kram (1981) showed many years later, the critical thing (on VI schedules with truly random distributions of reinforcers) is the *time since the last response* on each side. The longer the time since the animal last responded on a given side, the higher the probability that a response will be reinforced. You can see the truth of this in the concurrent FI FI case: on both sides, the longer the postreinforcement time, the more likely a response will be reinforced. Post*reinforcement* time is the critical thing here, because the FI timers run from one food delivery to the next. But post*response* time is the critical thing on random-interval (RI) schedules because of a curious property of randomness.

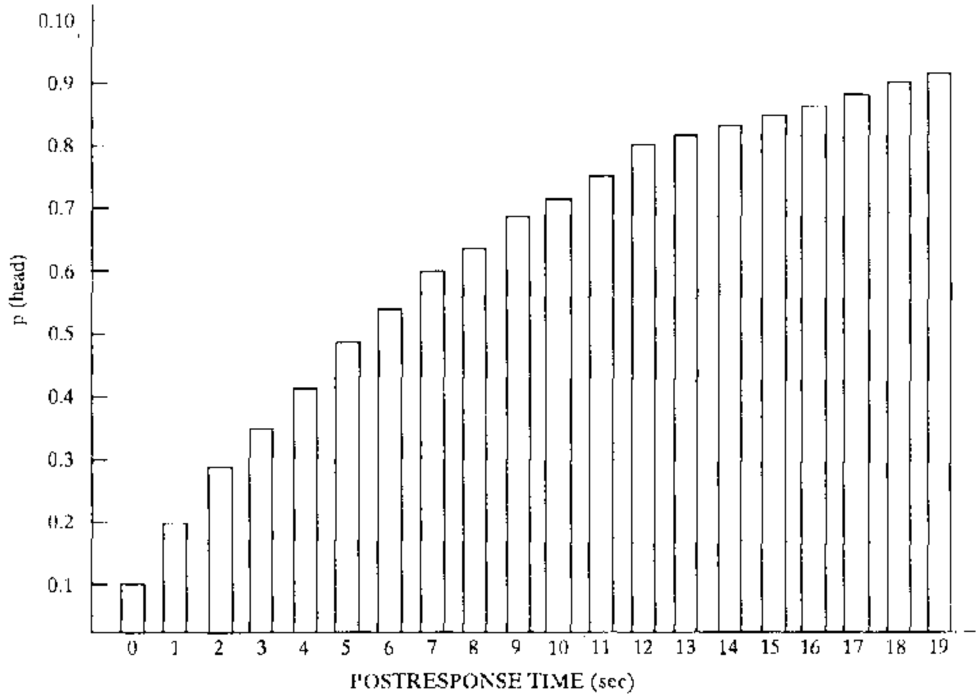
People get confused about this point, because of something called the gambler's fallacy. Suppose someone is tossing a coin, and you have been assured that it is a truly unbiased coin. Nevertheless, you have just witnessed 5 Heads in a row: what is the probability the next toss will be Tails? Obviously the answer depends in part on how much you trust the person who assured you the coin is unbiased. But trust us: it *is* unbiased. So what is the probability of a Tails on the next toss? Many people (probably not you, gentle reader) will guess that the probability of Tails is now *greater* than one-half, because of the prior preponderance of Heads. This is the *gambler's fallacy*, because (of course) the probability of Tails really is one half, just as you were told.

The counterpart to the gambler's fallacy with random-interval schedules is to believe that reinforcement is more likely the longer the time that has elapsed without reinforcement. This is not true, *so long as the animal continues to respond*. It *is* true, so long as he *does not* respond. Let's see why. On a random-interval schedule, reinforcement availability is determined by a process very like a sequence of coin tosses. The difference is that the tosses are much more likely to turn up Tails (no food) than Heads (food) and they are made very rapidly by a computer. And, most important of all, on interval schedules of all kinds, once the timer has timed out it does not start again until a response occurs: once a Head turns up, reinforcement is available for the next response, whenever it occurs.

Let's suppose the scheduling computer tosses its imaginary coin once per second, and the probability of Heads is 1/10. Then, so long as the animal responds at least once per second, he will get food on the average once every 10 seconds, i.e., a random-interval (RI) 10-s schedule. Moreover, if he responds exactly once every second, the probability each peck will result in food is a constant 1/10. But suppose he is interrupted, and *doesn't* respond, how will the probability of payoff change? Well, even if you're not a student of probability theory, you can see what will happen: The first second, the payoff probability is 1/10, but if he doesn't respond, then after two seconds, the payoff probability has gone up because Heads could have come up in the first second (he doesn't know, because he didn't respond), or the second second. The situation is like tossing a coin twice and asking: What is the probability of getting Heads at least once?

The algebra is very simple. The probability of at least one Head, in two tosses, is just one minus the probability of *two* Tails. For our RI schedule, the probability of a Tail is 9/10; the probability of getting a Tail on the second toss is also 9/10. Hence the probability of getting two Tails is $(9/10)_2 = 0.81$. Thus, the probability of at least one Head is $1 - 0.81 = 0.19$, a considerable increase over 0.1. The increment in p(Head) as a function of *postresponse time* is shown graphically in Figure 36. As you can see, after 20 seconds have gone by without a response, the probability of payoff is quite high, almost 0.9.

Figure 36



Probability of at least one Head after 1 to 20 tosses of a biased coin in which p(Heads) = 0.1 and p(Tails) = 0.9.

On a concurrent random-interval schedule we have two schedules like the one shown in Figure 36. On each key, therefore, the probability of reinforcement is rising, so long as the animal does not respond. As soon as he does, the probability goes back to zero and starts up the same function again. The difference between a rich RI schedule and a poor one is in the probability setting of the "coin tosser": the higher the probability of a Head, the faster the curve rises to its maximum value of one. If p(Head) = 1/2, for example, then p(Head) after one, two, three, etc. seconds of nonresponding is 0.5, 0.75, 0.875, etc.

The hypothesis of momentary maximizing says that *whenever* the organism decides to respond, it will always pick the alternative with the highest momentary probability of payoff, where probability of payoff is determined by a rising function like the one shown in Figure 36. Notice that the hypothesis is incomplete in an important way: it says which key will be chosen, once the animal decides to respond. But it does not say when that will be. That is, it doesn't say how fast the animal will respond or how its responses will be spaced in time. Nevertheless, it turns out that there is a simple way to see whether an animal is responding according to this rule. Suppose the two VI (really *RI*, random-interval) schedules are 1-min (on the Right) and 3-min (on the Left), for example. Then when he responds on the Left (the bad side), the time since his previous Left response should be at least three times as long as the time since his previous Right response. Conversely, when he responds on the Right, the time since his last Right response should be at least one third the time since his last Left response. Under most conditions following this rule implies that the animal will make three times as many responses on the Right as on the Left. This relationship should sound familiar: it is the ubiquitous matching result once again. In other words, MM, like essentially every other reinforcement rule, predicts matching in the concurrent RI RI situation.

What do the data say? We know that pigeons and rats match: do they match because they are also momentary maximizing? The answer seems to be "sometimes, but perhaps not every time". Hinson and Staddon (1983a & b) did a series of experiments in which they measured the times of every key peck in a concurrent RI RI situation. They found that pigeons follow the MM rule, but not terribly well. In their study, as in many others, the animals tend to *undermatch*, that is, the proportion of Left and Right pecks is closer to 50:50 (indifference) than the proportion of reinforcers.

Melioration

Howard Rachlin (1973) and Richard Herrnstein (Herrnstein & Vaughan, 1980) have suggested an alternative mechanism for matching that is somewhere in between the purely molecular mechanism of momentary maximizing and the purely molar relative law of effect (see Staddon, 1988, for a formal statement of the theory). Herrnstein calls this idea *melioration*, meaning "to make better". The closest analogy, for those of you familiar with elementary chemistry, is to the idea of *diffusion*. You may remember that if you have a concentrated solution on one side of a semipermeable membrane and a dilute solution on the other, water will diffuse from the more dilute into the more concentrated, because of osmotic pressure. The concentrated solution is for Herrnstein and Rachlin like the richer of two reinforcement schedules, and the flow of solute is like the shift of responding. Thus, melioration is a shift in preference towards the alternative in which the "cost" of reinforcement, i.e., responses made divided by reinforcements obtained, is less.

Note that this theory is similar to MM, in that the driving factor, the "cost" of reinforcement, is just the reciprocal of reinforcement *probability*. But it is different in that it does not specify over what time period the shift in preference should take place. In short, it is really a molar theory, because it contains unspecified assumptions about an averaging window over which reinforcement probability and response proportions are assessed (To be fair, MM is also silent about the averaging window within which payoff probability is measured).

What is the evidence for melioration? The strongest evidence is just that it so obviously gives the right answer: if preference shifts so that the cost of reinforcement is equal on both sides, i.e., $R(x)/x = R(y)/x$, then we automatically have matching. But of course it cannot explain systematic deviations, like the ubiquitous undermatching that is found in most experiments unless explicit steps are taken to prevent it (we discuss those steps in a moment). It is supported by the results of Lea's titration experiment, which i described earlier. But it has failed in one direct test.

The test used what is called a *frequency-dependent* (FD) schedule (Staddon, 1988). The FD schedule is like the two-armed bandit I have just discussed. The pigeon has two keys to peck, and payoffs for each choice are delivered probabilistically. The difference is that the payoff probabilities *depend on the animal's current preference*. Current preference is measured by the controlling computer, which keeps a record of the last 32 (say) choices. The FD schedule used by Horner and Staddon (1987) worked in the following way. Payoff probability on the Left varied from close to zero, to 0.2 (say). It was zero if the animal's current preference was 100% for the Left; it was 0.2 if his current preference was 100% for the Right, and proportionately less for intermediate values. Payoff probability for the Right followed a similar, linear function: zero if preference was 100% for the Left, but 0.1 if preference was 100% for the Right, so that no matter what the animal's current preference, payoff probability was always twice as high on the Left.

Notice that this FD schedule poses a particularly acute problem for a hill-climbing animal. If such an animal always picks the alternative with the higher payoff probability, then he must always pick the Left. But by doing so, he drives down the payoff probability ever lower. Indeed, when preference is exclusively Left, payoff probability is zero for both responses. A meliorating animal should therefore extinguish completely. A momentary maximizing animal is doomed to the same fate.

The brute fact is that Horner and Staddon's pigeons did *not* extinguish. They showed some preference for the Left, but it fell far short of the exclusive choice predicted by the two hill-climbing theories. Evidently, the birds were not blindly following either melioration or momentary maximizing. Horner and Staddon proposed an alternative, a probabilistic learning rule they call *ratio invariance*, that is consistent with their results but it is too technical to pursue here. For the moment, it is sufficient to say that neither melioration nor momentary maximizing are adequate to account for probabilistic choice.

Linear Waiting

Earlier i showed how proportional timing seems to underlie a wide variety of effects on chain reinforcement schedules. Since this is such a ubiquitous process, we need to see what it predicts for the situations now under discussion. What does proportional timing imply for molar measures of performance on RI and FR schedules, for example? What does it imply for the allocation of behavior on concurrent RI RI? Is it consistent with the matching law? I will take these cases in turn.

Simple RI Schedules. Behavior on RI schedules is the easiest to understand in terms of proportional timing. The first thing we need to do is to decide on the time marker. In the discussion of momentary maximizing i showed that postresponse time is the critical variable on RI schedules.

Proportional timing then just says that pigeons (and rats) will set their average interresponse time to be a fixed proportion of the expected time between a response and reinforcement. If reinforcement is delivered on a random basis, then the expected time to reinforcement will just be 1/reinforcement rate. Thus, proportional timing implies that response rate should be proportional to reinforcement rate on RI schedules. This is close to the truth when reinforcement rate is low; it is not true when reinforcement rate is high. What have we forgotten?

We have forgotten that there is an upper limit on response rate, a lower limit on the time between responses. When typical waiting times are long, on FI schedules for example, we can ignore this constraint. But when they are short, on RI schedules for example, we need to include it. If we add in a term for the minimum time between responses, we have a modified version of proportional timing that Wynne and Staddon (1988) have called *linear waiting*. If we denote the expected time to reinforcement by I and the time between responses by t, linear waiting is just

$$t = AI + B, \tag{12}$$

where A is the proportion of the expected time to food that the animal waits, as before, and B is the shortest possible interresponse time. Thus, even if I is zero, the animal will wait at least time B between successive responses.

Now let's look at what Equation 12 predicts for simple RI schedules. t is just 1/response rate, x; I is just 1/reinforcement rate, R(x). Making these substitutions yields

$$\begin{aligned} 1/x &= A/R(x) + B, \text{ or } x = R(x)/[A + BR(x)] \\ &= (1/B)R(x)/[A/B + R(x)]. \end{aligned} \tag{13}$$

Equation 13 should be very familiar: it is just Herrnstein's hyperbola, the matching law prediction for simple VI schedules (Equation 11), with $1/B = k$, and $A/B = R_0$. As we have seen, this equation gives a good description of molar responding on VI schedules (cf. Figure 18). Moreover, the parameters 1/B and A/B have the same significance as Herrnstein's parameters k and R_0. Recall that R_0 was supposed to represent the effect of "other" reinforcers. Hence, increasing the magnitude of the reinforcer for response X should in effect reduce R_0. Increasing reinforcement magnitude reduces the waiting proportion, A, as we saw earlier, hence reduces A/B. Parameter k is assumed to represent the maximum possible response rate;

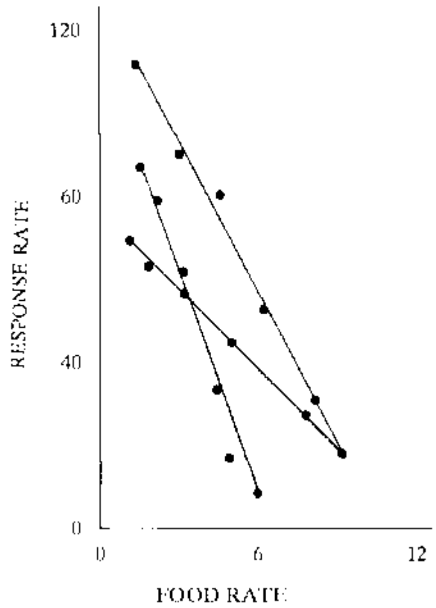
the linear-waiting parameter B represents the shortest possible interresponse time, so that 1/B has the same significance as k.

Fixed-Ratio Schedules. One of the main problems with the single-response matching equation (Equation 11) is that it does not give an accurate prediction of how response rate varies with reinforcement rate on fixed-ratio schedules (cf. Timberlake, 1977). And since it has no term for the schedule itself, it obviously cannot describe the relation between response rate and ratio value. It is possible to derive these predictions from linear waiting, however.

Earlier I argued that fixed-ratio schedules are treated by animals in a fashion very similar to two-link chain schedules. We can use this analysis to work out how overall response rate should depend on ratio value, and how it should be related to obtained reinforcement rate as the ratio value is varied (i.e., the response function). The algebra involved is too lengthy to go over here, but the conclusions are quite simple: (a) Linear waiting predicts a positive relation between ratio value and response rate. (b) Linear waiting predicts a linear relation, with negative slope, between response rate and obtained reinforcement rate. The validity of these predictions depends on conditions being such that the parameters A and B are constant. They will not be constant if the animal's motivational state is allowed to vary if the experimental session is so long, or the reinforcement rate so high, that he is more or less hungry at the end of the session than at the beginning, for example.

You have already seen data consistent with the first prediction in Figure 22: the results of the Hirsch and Collier experiment, as well as numerous others reviewed by Hogan and Roper (1978), show that response rate increases with ratio value over most of the range in most experiments with food or water reinforcement. The second prediction, a linear response function of negative slope, has also been repeatedly confirmed: in ratio-schedule experiments by Allison that are not open to the artifactual objection we discussed earlier (see Allison, 1983), and by Ettinger and Staddon (1983), using a novel procedure in which rats working for food reinforcement were repeatedly exposed to an ascending and descending cycle of six ratio values, (2,4,8,16,32 and 64). Rats adapt quickly to this procedure, and respond at a rate and with a postreinforcement pause appropriate to the prevailing ratio. The virtue of the cyclic procedure, therefore, is that it allows us to measure the effect of ratio value on responding under conditions where everything else is held constant this is not always possible when comparisons must be made between one block of daily experimental sessions and another. By comparing cycles early and late in the session we can make sure that performance is the same at the beginning as at the end, permitting the six cycles/session to be averaged.

Figure 37



Each point is the average lever press rate for a group of four rats' trained on a cyclic- ratio schedule. Each point is for a different ratio value: the points at the ottom of the graph, for example, are for FR 2, the points at the top for FR 64. The straight lines are fitted to the data. The rightmost line is for rats at 80% of their free-feeding weights, the inner line parallel to it is for the same rats getting food adulterated with quinine. The line of shallow slope is for the rats at 95% of their free-feeding weights (from Ettinger & Staddon, 1983).

The results are shown in Figure 37. The figure shows three straight-line response functions. The one to the right was obtained with very hungry rats (80% of free-feeding weight), the parallel one to the left was obtained with the same rats eating food adulterated with bad-tasting quinine. The third function, with lower slope, was obtained when the rats were less hungry (95% of free-feeding weight). What will be the effect of hunger, in terms of the linear waiting model? Recall that earlier I used the fact that the waiting fraction, A, is smaller for large reinforcements to explain some properties of self control on chain schedules. We can do something similar here, because the reinforcer must change in effective value as the animal becomes satiated. Hence, parameter A should be larger for a satiated animal than for a hungry one. The algebra shows that an increase in A means a shallower response-function line which is just what we see in the Figure. Now what about the y-intercept? It also changes when A changes, and Ettinger and Staddon found that as well (see their Figure 8). But notice that anything that affects only parameter B affects *only* the y-intercept, and not the slope. Apparently a noxious taste, such as quinine, does just that, as shown by the leftmost parallel response function in Figure 37. Linear waiting cannot be the whole story here, however, because it predicts a too-shallow slope for the typical FR response function. Regulatory processes, such as the one suggested by Ettinger & Staddon seem also to be involved.

Concurrent RI RI Schedules. We have already seen that any reinforcement rule that satisfies two not-very-restrictive conditions that response rate be positively related to reinforcement rate, with zero intercept is sufficient to produce matching on concurrent RI RI schedules. Clearly, linear waiting satisfies these conditions: interresponse time is positively related to reinforcement delay, and infinite delay (zero reinforcement rate) should produce infinite waiting (no responses). Hence, linear waiting, like numerous other reinforcement rules, predicts matching of rsponse ratios to ratios of obtained reinforcement on concurrent RI RI.

We have also seen that behavior on concurrent RI RI often conforms quite well to momentary maximizing: does this result conflict with the linear-waiting hypothesis? Not at all, because an animal that follows the linear-waiting rule will also show momentary maximizing. For example, suppose the RI schedule on the Left is 3-min and on the Right 1-min. Linear waiting implies that the time between interresponse times on the Left will be on average about three times longer than IRTs on the Right. Such a pattern is not *required* by MM recall that the actual times when a "decision" to respond is made are not prescribed by MM -but it is perfectly consistent with MM, because the response that actually occurs will satisfy the MM condition, that it is made to the alternative with highest payoff probability.

Undermatching is a frequent deviation from simple matching. *Undermatching* means that the ratio of responses, x/y, is closer to indifference (50:50) than the ratio of reinforcements. For example, an animal that shows x/y = 2.5 and R(x)/R(y) = 3 is said to undermatch. What is the reason for this deviation? Linear waiting depends on the animal's ability to remember the time marker, that is, to remember where he responded last, and to remember which response he was making when reinforcement occurred.

Anything that impairs recency discrimination will cause confusions between one choice alternative and the other, hence will make the alternatives more similar, hence will favor undermatching. Conversely, anything that makes the two alternatives more distinct should shift preference in the opposite direction, towards (but not beyond) matching.

Sheer lapse of time impairs memory. Hence, a concurrent schedule in which reinforcement is infrequent, which implies long waiting times, should cause more memory confusions. Are animals more likely to show undermatching on lean concurrent schedules? The answer seems to be "yes". For example, Fantino, Squires, Delbrück and Peterson (1972) looked at matching in pigeons responding for food reinforcement on different pairs of VI schedules ranging from 6 s vs. 12 s to 600 s to 1200 s. They found responding to each key was almost equal at the longest values, whereas choice was almost exclusive at the shortest. The undermatching was also greater at the longer values, but the effect was relatively small. Because of the properties of the VI feedback function I discussed earlier, even close-to-exclusive choice nevertheless produces pretty good matching, so a tendency to confuse the two keys is partially compensated for by comparable changes in the relative frequency of reinforcements obtained: as responding declines in absolute level, as preference approaches indifference there is a strong tendency for obtained reinforcement rates to equalize as well. Memory for an event is also impaired if it occurs close in time to other, similar events. If animals switch frequently between choice alternatives, they may have difficulty remembering which choice they made last. A procedure known as the *changeover delay* (COD) was devised early on as a way to guarantee matching on concurrent RI RI schedules. The COD also discourages animals from switching frequently between alternatives. It works by imposing a "dead time" of a second or two after each switch; during this COD time, no reinforcement can be obtained, even if the RI timer has set one up. Animals trained in this way soon learn not to switch frequently between alternatives, and of course when reinforcement does occur it is separated by at least the COD time from responding to the nonreinforced alternative. Both the lowered switching rate and the separation in time of reinforcement for the two choices aid memory and so should favor matching over undermatching.

Molecular Mechanisms: Conclusion

I have discussed three molecular mechanisms for behavioral allocation: momentary maximizing, melioration, and linear waiting. The first two are hill-climbing processes; they act to shift preference in the direction of the locally better alternative. Both therefore predict fixation on the richer alternative in the two-armed bandit situation. There is also some evidence for MM on concurrent RI RI schedules. Melioration explains matching on concurrent RI RI and RI RR reasonably well, but must bring in auxiliary assumptions to explain undermatching, a common deviation from matching. Melioration, in the form of the relative law of effect, can account for molar behavior on simple RI schedules, but fails to account for behavior on simple FR schedules because it treats all schedules alike. Momentary maximizing is also unable to account for performance on simple schedules. Neither melioration nor matching can explain performance on a frequency-dependent schedule in which the probability of payoff on the one side is always higher than on the other, and increasing preference for the richer side reduces the absolute values of both payoff probabilities.

The timing process called linear waiting is able to account for molar patterns on simple schedules. It explains why the hyperbolic equation relating response and reinforcement rates on RI schedules works so well, and it can also explain the very different response function on FR schedules. On concurrent RI RI schedules, linear waiting is very similar to momentary maximizing and makes similar predictions. It can also account for the effects of the changeover delay, which tends to produce good matching, and the effects of low absolute reinforcement rates, which tend to produce undermatching. We do not know (it is not easy to solve the problem analytically) whether linear waiting can account for the complex patterns of behavior produced by some frequency-dependent schedules, however, nor is it clear whether it can account for biased matching on concurrent VI VR schedules.

References

Allison, J. (1980). Conservation, matching, and the variable-interval schedule. *Animal Learning & Behavior, 8*, 185-92.

Allison, J. (1983). *Behavioral economics.* New York: Praeger.

Atkinson, J. W., & Birch, D. (1970). *The dynamics of action.* New York: Wiley.

Bernoulli, D. (1738). Exposition of a new theory on the measurement of risk. Reprinted in G. A. Miller (Ed.). *Mathematics and Psychology.* 1964.

Catania, A. C., & Reynolds, G. S. (1968). A quantitative analysis of the behavior maintained by interval schedules of reinforcement. *Journal of the Experimental Analysis of Behavior, 11*, 327-383.

Catania, A. C. (1970). Reinforcement schedules and psychophysical judgements. In W. N. Schoenfeld (Ed). *The theory of reinforcement schedules.* New York: Appleton-Century Crofts.

Catania, A. C., Matthews, T. J., Silverman, P. J., & Yohalem, R. (1977). Yoked variable-ratio and variable-interval responding in pigeons. *Journal of the Experimental Analysis of Behavior, 28*, 155-161.

Catania, A. C. (1979). *Learning*. Englewood Cliffs: Prentice Hall.

Ettinger, R. H., & Staddon, J. E. R. (1983). Operant regulation of feeding: a static analysis. *Behavioral Neuroscience, 97*, 639-653.

Ettinger, R. H., Reid, A. K., & Staddon, J. E. R. (1987). Sensitivity to molar feedback functions: a test of molar optimality theory. *Journal of Experimental Psychology: Animal Behavior Processes, 13*, 366-375.

Fantino, E. (1969). Choice and rate of reinforcement. *Journal of the Experimental Analysis of Behavior, 12*, 723-730.

Fantino, E., Squires, N., Delbruck, N., & Peterson, C. (1972). Choice behavior and the accessibility of the reinforcer. *Journal of the Experimental Analysis of Behavior, 18*, 35-44.

Fantino, E. (1977). Conditioned reinforcement. In W. K. Honig, & J. E. R. Staddon (Eds.) *Handbook of operant behavior.* Englewood Cliffs, NJ: Prentice-Hall.

Fantino, E., & Abarca, N. (1985). Choice, optimal foraging and the delay reduction hypothesis. *Behavioral and Brain Sciences, 8*, 351-430.

Gibbon, J. (1977). Scalar expectancy and Weber's law in animal timing. *Psychological Review, 84*, 279-325.

Gibbon. J.. Church, R. M., Fairhurst, S., & Kacelnik, A. (1988). Scalar expectancy theory and choice between delayed rewards. *Psychological Review, 95*, 102-104.

Green, L., & Snyderman, M. (1980). Choice between rewards differing in amount and delay: Toward a choice model of self control. *Journal of the Experimental Analysis of Behavior, 34*, 135 147.

Green, L., Kagel, J. H., & Battalio, R. C. (1982) Ratio schedules of reinforcement and their relation to economic theories of labor supply. In M. L. Commons, R. J. Herrnstein, & H. Rachlin (Eds.) *Quantitative analyses of behavior. Vol II. Matching and maximizing accounts.* Cambridge, MA: Ballinger.

Herrnstein, R. J. (1961) Relative and absolute strength of response as a function of frequency of reinforcement. Journal of the Experimental Analysis of Behavior, *4*, 267-272.

Herrnstein, R. J. (1964). Secondary reinforcement and rate of primary reinforcement. *Journal of the Experimental Analysis of Behavior, 7*, 27-36.

Herrnstein, R. J. (1966). Superstition: A corollary of the principles of operant conditioning. In W. K. Honig (Ed.) *Operant behavior* New York: Appleton Century-Crofts. pp. 33-51.

Herrnstein, R. J. (1970). On the law of effect. *Journal of the Experimental Analysis of Behavior, 13*, 243 266.

Herrnstein, R. J., & Loveland, D. H. (1975). Maximizing and matching on concurrent ratio schedules. *Journal of the Experimental Analysis of Behavior, 24*, 107 116.

Herrnstein, R. J., & Heyman, G. M. (1979). Is matching compatible with reinforcement maximization on concurrent variable interval variable ratio? *Journal of the Experimental Analysis of Behavior, 31*, 209-223.

Herrnstein, R. J., & Vaughan, W. (1980). Melioration and behavioral allocation. In J. E. R. Staddon (Ed.). *Limits to Action: The Allocation of Individual Behavior*. New York: Academic Press, pp. 143-176.

Hinson, J. M., & Staddon, J. E. R. (1978). Behavioral competition: A mechanism for schedule interactions. *Science, 202*, 432-434.

Hinson, J. M., & Staddon, J. E. R. (1983a). Hill-climbing by pigeons. *Journal of the Experimental Analysis of Behavior, 39*, 25-47.

Hinson, J. M., & Staddon, J. E. R. (1983b). Matching, maximizing and hill climbing. *Journal of the Experimental Analysis of Behavior, 40*, 321-31.

Hirsch, E., & Collier, G. (1974). Effort as determinant of intake and patterns of drinking in the Guinea pig. *Physiology and Behavior, 12*, 647-655.

Hogan, J. A., & Roper, T. J. (1978). A comparison of the properties of different reinforcers. In: *Advances in the Study of Behavior* (J. S. Rosenblatt, R. A. Hinde , E. Shaw & C. Beer), (Eds.) Vol. 8. New York: Academic Press.

Horner, J. M., & Staddon, J. E. R. (1987). Probabilistic choice: A simple invariance. *Behavioural Processes, 15*, 59-92.

Hursh, S. R. (1984). Behavioral economics. *Journal of the Experimental Analysis of Behavior, 42*, 435-452.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*, 263-291.

Kahneman, D., & Tversky, A. (1984). Choices, values and frames. *American Psychologist, 39*, 341-350.

Kelleher, R. T. (1966). Conditioned reinforcement in second-order schedules. *Journal of the Experimental Analysis of Behavior, 9*, 475-486.

Killeen, P. (1985). Incentive theory: IV. Magnitude of reward. *Journal of the Experimental Analysis of Behavior*, *43*, 407-417.

Krebs, J. R., Erichsen, J. T., Webber, M. I., & Charnov, E. L. (1977). Optimal prey selection by the great tit (*Parus major*). *Animal Behaviour*, *25*, 30-38.

Lea, S. E. G. (1978). The psychology and economics of demand. *Psychological Bulletin*, *85*, 441-446.

Lea, S. E. G., & Tarpy, R. M. (1982). Different demand curves from rats working under ratio and interval schedules. *Behaviour Analysis Letters*, *2*, 113-121.

Le Magnen, J. (1985) *Hunger*. Cambridge University Press.

Lorenz, K. (1952). The past twelve years in the comparative study of behavior. In C. H. Schiller (Ed.). *Instinctive Behavior*, New York: International Universities Press.

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. New York: Wiley.

Luce, R. D. (1977). The choice axiom after twenty years. *Journal of Mathematical Psychology*, *15*, 215-233.

MacArthur, R. H., & Pianka, E. R. (1966). On optimal use of a patchy environment. *American Naturalist*, *100*, 603-609.

MacEwen, D. (1972). The effects of terminal-link fixed-interval and variable-interval schedules on responding under concurrent chained schedules. *Journal of the Experimental Analysis of Behavior*, *18*, 253-261.

Mazur, J. E., Snyderman, M., & Coe, D. (1985). Influences of delay and rate of reinforcement rate as determinants of schedule preference. *Journal of Experimental Psychology: Animal Behavior Processes*, *11*, 565-575.

Mazur, J. (1986). Choice between single and multiple delayed reinforcers. *Journal of the Experimental Analysis of Behavior*, *46*, 67-78.

Mook, D. G. (1987). *Motivation: The organization of action*. New York: Norton.

Moore, J. (1982). Choice and number of reinforcers. *Journal of the Experimental Analysis of Behavior*, *37*, 115-122.

Neuringer, A. J., & Chung, S. H. (1967). Quasi-reinforcement: control of responding by a percentage-reinforcement schedule. *Journal of the Experimental Analysis of Behavior*, *10*, 45-54.

Powell, R. W. (1969). The effect of reinforcement magnitude upon responding under fixed-ratio schedules. *Journal of the Experimental Analysis of Behavior*, *12*, 605-608.

Premack, D. (1965). Reinforcement theory. In D. Levine (Ed.), *Nebraska Symposiumon Motivation* (Vol. 13). Lincoln: University of Nebraska Press.

Rachlin, H. (1973). Contrast and matching. *Psychological Review, 80,* 217-234.

Rachlin, H. (1978). A molar theory of reinforcement schedules. *Journal of the Experimental Analysis of Behavior, 30,* 345-360.

Rachlin, H., Battalio, R. C., Kagel, J. H. & Green, L. (1981). Maximization theory in behavioral psychology. *The Behavioral and Brain Sciences, 4,* 371-388.

Rescorla, R. A. (1982). Comments on a technique for assessing associative learning. In M. L. Commons, R. J. Herrnstein, A. R. Wagner. *Quatitative Analyses of Behavior. Acquisition.* Vol. III. Cambridge Massachusetts: Harper & Row.

Samuelson, P. A. (1965) *Foundations of economic analysis.* Cambridge, Mass.: Harvard University Press.

Shimp, C. P. (1966). Probabilistically reinforced choice behavior in pigeons. *Journal of the Experimental Analysis of Behavior, 9,* 443-455.

Shull, R. L., & Spear, D. J. (1987). Detention time after reinforcement: effects due to delay of reinforcement? In M. L. Commons, J. E. Mazur, J. A. Nevin & H. Rachlin (Eds.), *Quantitative analyses of behavior: Vol. 5. The effect of delay and intervening events on reinforcement value* (pp. 187-204). Hillsdale, NJ: Erlbaum.

Smith, A. (1976). *An Inquiry into the Nature and Causes of the Wealth of Nations Oxford University Press.* (reprint of 1776 edition).

Squires, N., Norborg, J., & Fantino, E. (1975). Second-order schedules: discrimination of components. *Journal of the Experimental Analysis of Behavior, 24,* 157-171.

Staddon, J. E. R. (1972). Temporal control and the theory of reinforcement schedules. In R. M. Gilbert & J. E. Millenson (Eds.). *Reinforcement: Behavioral Analyses.* New York: Academic Press. (pp. 209-262).

Staddon, J. E. R. (1974). Temporal control, attention and memory. *Psychological Review, 81,* 375-391.

Staddon, J. E. R. (1976). Operant behavior as adaptation to constraint. Contribution to Symposium on Response Strength (with C. Shimp, A. Nevin, R.J. Herrnstein), *American Psychological Association Meeting,* Washington, D.C.,September, 1976.

Staddon, J. E. R., & Motheral, S. (1978). On matching and maximizing in operant choice experiments. *Psychological Review, 85,* 436-444.

Staddon, J. E. R. (1979). Operant behavior as adaptation to constraint. *Journal of Experimental Psychology: General, 108,* 48-67.

Staddon, J. E. R. (1980). Optimality analyses of operant behavior and their relation to optimal foraging. In J. E. R. Staddon (Ed.), *Limits to action: The allocation of individual behavior.* New York: Academic Press.

Staddon, J. E. R., Hinson, J. M., & Kram, R. (1981). Optimal choice. *Journal of the Experimental Analysis of Behavior, 35,* 397-412.

Staddon, J. E. R., & Hinson, J. M. (1983). Optimization: A result or a mechanism? *Science, 221,* 976-7.

Staddon, J. E. R. (1988). Quasi-dynamic choice models: melioration and ratio-invariance. *Journal of the Experimental Analysis of Behavior, 49.* 303-320.

Stephens, D. W., & Krebs, J. R. (1986). *Foraging theory.* Princeton University Press.

Stubbs, D. A. (1971). Second-order schedules and the problem of conditioned reinforcement. *Journal of the Experimental Analysis of Behavior, 16,* 289-313.

Timberlake, W., & Allison, J. (1974). Response deprivation: An empirical approach to instrumental performance. *Psychological Review, 81,* 146 164.

Timberlake, W. (1977). The application of the matching law to simple ratio schedules. *Journal of the Experimental Analysis of Behavior, 25,* 215-217.

Timberlake, W., & Peden, B. F. (1987). On the distinction between open and closed economies. *Journal of the Experimental Analysis of Behavior, 48,* 35-60.

Werner, E. E., & Hall, D. J. (1974). Optimal foraging and the size selection of prey by the bluegill sunfish (*Lepomis macrochirus*). *Ecology, 55,* 1042-1052.

Williams, B. A. (1979). Contrast, component duration, and the following schedule of reinforcement. *Journal of Experimental Psychology: Animal Behavior Processes, 5,* 379 396.

Wynne, C. D. L., & Staddon, J. E. R. (1988). Typical delay determines waiting time on periodic- food schedules: static and dynamic tests. *Journal of the Experimental Analysis of Behavior, 50,* 197-210.

Box 1: Proportional Timing and Optimal Policy in Self-Control Experiments

We assume that the animal treats each key independently, and that the initial-key VI can be treated as an FI, but with a smaller proportional-waiting constant, a. Thus, for the procedure in Figure 12:

pause on the LEFT: $a_L(t+T)$ (B9.2.1A)

on the RIGHT: $a_R(t+KT)$, (B9.2.1B)

where K is the ratio of Long/Short second links. We also assume that the pause constants a_L and a_R will be inversely related to the reward magnitudes on the Left and Right. If the ratio $a_L/a_R = A$, and we rewrite a_R as just a, for simplicity, eq. B9.2.1 reduces to

pause on the LEFT: $Aa(t+T)$ (B9.2.2A)

on the RIGHT: $a(t+KT)$, (B9.2.2B)

Response rate, by our assumption (and ignoring competition) in the first link is proportional to the time spent responding on each side, which is just t minus pause time. The ratio of responding is just the ratio of these quantities:

LEFT/RIGHT $= [t-Aa(t+T)]/[t-a(t+KT)] = [t(1-Aa)-aAT]/[t(1-a)-aTK]$. (B9.2.3)

(Competition will just tend to make preferences more extreme, because it favors the side with the shorter pause, but affects both sides equally when pauses are equal). Equation B9.2.3, with the indicated parameter values, generated the curves in Figures 9.13–9.15. Note that if $T = 0$, eq. B9.2.3 reduces to $L/R = (1-Aa)/(1-a)$, which is unity if the two rewards are equal, and shows a bias towards the large-reward side if $A>1$.

Optimal policy in the self-control procedure. There are three possible strategies here: *impulsive* (always pick the short-delay side), self-*control* (always pick the long-delay side) and *sample pick both*. The *sample* strategy is only a possibility with VI initial links. For the simple FI case the comparison between *impulsive* and *control* is straightforward. The food rates are

impulsive: $1/(t+T)$

control: $A/(t+KT)$, $A,K >= 1$.

where A is the ratio of large/small reward magnitude and K the ratio of long/short second-link length. These two yield the switching condition, choose *impulsive* if

$t < T(K-A)/(A-1)$, or $t/T < (K-A)/(A-1)$. (B9.2.4)

Note that the choice here does not depend on the absolute values of T and t, but only on their ratio, which is different both from the data and the proportional-timing prediction. Moreover, if $K<=A$, the animal should never choose *impulsive*, and if $A<=1$, he should always choose impulsive.

The prediction with VI first links is too complicated to go into here. The food rate for *sample* is given by a generalized version of the argument we gave earlier for chain VI FI with identical rewards on both sides. In the present case it yields $A/(t+2KT)+1/(t+2T)$ for the sample food rate, which in turn gives rise to quadratic expressions when compared with the food rates for the other two strategies.

Box 2: The Laffer Curve

The "Laffer curve" was reportedly sketched by California economist Arthur Laffer on a paper napkin as he vigorously argued the case for lower tax rates to a colleague over lunch. He justified his proposal by the following argument. Suppose that people's willingness to do productive work depends upon the effective wage rate in the backward-bending fashion of the standard labor-supply curve. The effective wage rate depends on the tax rate, since we can only spend what is left after income tax has been extracted. Tax rates are (nominally) highly progressive, so that the higher the wage rate the higher the proportion that is lost (to the earner) in taxes. Hence, there must be a point at which higher taxes begin to depress the total amount of productive work that is done. If the total amount of work done decreases enough, then the total tax "take" will begin to decrease, even though the tax rate is high.

The key question is: When is the tax rate too high? That is, when is it so high that the depressive effect on total amount of labor supplied exceeds the added revenue brought in by virtue of the high rate? Laffer argued that time is *now*, so that taxes ought to be reduced.

Laffer's argument appears to be an economic one, but it clearly has a very large, and crucial, psychological component, because it hinges on the size of the reduction in work associated with loss of income to taxes. If people remain willing to work hard, even though a large chunk of their income goes to taxes, then Laffer's argument fails. The current consensus seems to be that people are more willing to work than Laffer expected. Yet major legislation was enacted based on little more than a collective conjecture about how people would react to a small increase in take-home pay.