# 9.0 WHAT ARE THE EFFECTS OF REINFORCER PRESENTATIONS?

## 9.1 Questions Relating to Reinforcement

The term "reinforcer" refers to a type of stimulus event that produces certain behavioral effects in some situations and circumstances, some of the time. The need to use the word "some" twice in the above sentence, bears testimony to the fact that the concept of "reinforcer," though often used as if it were well understood, actually conceals many as-yet unanswered questions.

One of the important categories of variables on which the effects of stimuli termed "reinforcers" depend (besides the preceding behavior) are "establishing operations" (Michael, 1982). They depend also on the recency and recent density of similar stimulus events, on the organism's history relating to similar events in similar circumstances, on the history, type, and frequency of behavior on which the event impacts (Morse, 1966, pp. 54-55), on prevailing reinforcement contingencies and stimuli, and on the organism's level of arousal (Killeen, Hanson, & Osborne, 1978).

The term "reinforcer" itself is semi-descriptive, as it implies that the presentation of such a stimulus produces a type of behavioral effect that can be described as "reinforcement." But one of the frontiers of operant behavior research is the description and analysis of the various types of behavioral effects *actually* produced by the presentation of presumptive "reinforcers," and the independent variables on which those effects depend (Morse, 1966, p. 55). The main categories of such independent variables are set forth in the above paragraph. The likely reason why this area of research has remained unexplored is that it cannot be addressed by means of iOs. The sections that follow attempt to show how it can be addressed by means of rOs.

## 9.2 Do Reinforcer Presentations Affect Individual Occurrences of Operants?

An example of an open question regarding reinforcement is which, if any, behavioral effects of the presentation of a presumptive reinforcer (let's call that a "PPRf" for short) can be observed at the level of individual occurrences of operants.[1] A closely related question is what (if any) behavioral effects are produced by a single PPRf, and how these effects summate to produce the known effects of multiple PPRfs. Multiple PPRfs can result from a reinforcement contingency maintained for an extended time.

Since the traditional iO techniques do not provide a way to address these questions, some researchers have, from time to time, attempted to approach them by visually observing and comparing the topography of the behavior immediately preceding and following a PPRf (e.g.,

[1] It would be ungrateful of me not to mention here one of the unforgettably provocative comments that William N. Schoenfeld made in a graduate psychology seminar at Columbia University in 1952. "We don't even know the effects of a single reinforcement presentation on an individual response", he said, to drive home the point that learning theory is still in its infancy. The comment would be as valid today as it was then.

Muenzinger, 1928; Skinner & Morse, 1958; Iversen, 1982). Such visual comparisons have never established that the PPRf produced a repetition of the immediately preceding behavior.

This failure may have been due to the fact that every PPRf has a dual function: that of reinforcement and that of a discriminative stimulus ($S^D$) for some behavior that was previously reinforced in a similar situation. It is plausible that the effect due to the $S^D$ function normally overwhelms and obliterates the possible reinforcement effect, leaving unanswered the question of what (if any) "reinforcement" effect is *ever* present at the individual occurrence level.

A word of explanation may be in order regarding the important concept of the "$S^D$ function of a PPRf." Every PPRf is preceded by and is concurrent with certain other events. Those events include the subject's own pre-PPRf behavior and the recent density of PPRfs. For example, a long stretch of behavior without PPRfs can comprise part of the recent events. Thus, the PPRf is a compound stimulus with several identifiable components: (a) the PPRf itself, (b) the subject's immediately preceding behavior, (c) the schedule on which reinforcers have recently been received, and (d) the exteroceptive stimuli comprising the physical environment in the presence of which the PPRf is presented. That compound stimulus will always generalize, to some degree, with sets of circumstances that occurred earlier in the subject's history, perhaps minutes, hours, days, or months earlier. The degree of generalization will depend on the similarity of those sets of circumstances. One must therefore expect the compound stimulus inherent in any PPRf to act like an $S^D$, setting the occasion for whatever behavior was shaped by the contingencies that prevailed right after a previous occurrence of a similar compound stimulus and PPRf. To the degree that this happens, every PPRf functions as an $S^D$.

The rO technique provides a way to separate the $S^D$ effects from the other effects of PPRfs, and for observing and measuring the effects of PPRfs at the micro or molecular level of individual occurrences, rather than merely at the macro or molar statistical mass-action level. The rO also provides a way to investigate how these effects depend on the independent variables listed in Section 9.1.

## 9.3 Mechanisms of Shaping

There is also a question that transcends the molar versus molecular one: *How* do PPRfs shape operant behavior (Morse, 1966, p. 56)? That question is valid regardless of whether shaping operates at the individual occurrence level or only at the molar level. The "how" question calls for an explication of mechanisms.

The shaping process is explained in the literature by reference to the molar process of "successive approximations" and "response differentiation." According to this explanation, shaping occurs when reinforcers selectively impinge on the response variants that fall to the chosen side of the variability distribution for a chosen criterion, thereby progressively shifting that distribution in the desired direction by a cumulative statistical action (e.g., Skinner, 1938, pp. 312-338; Keller & Schoenfeld, 1950, pp. 164-190; Wilson & Keller, 1953; Herrick, 1964; Morse, 1966, p. 55; Pear & Legris, 1987).

Although that explanation is qualitatively consistent with much experimental data, it does not account persuasively for the remarkable speed and efficiency with which the shaping process often proceeds. Skilled animal trainers can shape behavior so fast that it sometimes seems almost as if they were telling the animal what to do. A skilled trainer clearly does not rely on the inherently slow progressive statistical shifting of variability distributions (Pierrel & Sherman, 1963). The explanation also leaves open the question of whether selection of response variants is indeed the mechanism that underlies shaping.[2]

To be satisfying, an explanation of the shaping process would have to describe (a) the mechanisms that generate the variants from which selection can take place, and (b) the proximal (molecular) behavioral effects of a selection event. The remainder of this chapter attempts to show how the rO permits such questions to be addressed. Section 9.6 below outlines a possible alternative mechanism for the shaping process, and Section 9.9 addresses issue (a) above. The difficulty of addressing these issues with iOs may be the reason why even an approach or strategy for addressing them has so far remained elusive.

## 9.4 Reinforcement as a Parameter Shifter

Research in the field of motor behavior, most of it done since 1975, teaches us that well-established operant behavior routines become linked and coordinated in ways that allow them to be flexibly specified (at the CNS level) by attribute parameters. Examples of important attribute parameters are response force (which generally corresponds to muscle potential or degree of muscle engagement, which in turn corresponds to placement along the dimension of overtness-covertness); the timing and phasing of the muscular contractions: and the particular system of effectors that produces the operant's defining effect (Stelmach, Mullins, & Teulings, 1984; Kelso, Tuller, Vatikiotis-Bateson, & Fowler, 1984; Summers, Sargent, & Hawkins, 1984; Rosenbaum, 1985; Ivry, 1986; Schmidt, 1988, pp. 187-298; Pew & Rosenbaum, 1988; Keele, Cohen, & Ivry, 1990; Semjen & Gottsdanker, 1990; Wiesendanger, 1990).[3] Thinking of operants as behavior routines modified by parameter settings suggests some possible mechanisms for the action of reinforcement.

One such mechanism is that a PPRf results in the repetition not of the most recent *behavior*, but rather of its most recent *direction of change*. In other words, the PPRf operates on the operants' parameter settings more like a vector than like a duplicator. It perpetuates its most recent *shift*, not its most recent setting. By way of an oversimplified illustration, if a certain operant has recently occurred twice, and if, in those two occurrences, the setting of one of its parameters shifted from 7 to 8, for whatever reason, a reinforcer presented after the second of those two occurrences would shift the setting again, this time from 8 to 9. (These numbers are only illustrative, of course). Presenting the reinforcer says, in effect, "Keep going in that direction," as one says in the children's parlor game "warmer/colder" in guiding the player toward a chosen

[2]Skinner's observation that selection operates both in the shaping of operant behavior and in the shaping of species did not require an explication of the mechanisms by which the selected behavioral variants are generated, any more than Darwin needed to explain the origins of biological variants.

[3]These references, several of which are reviews of the literature, constitute a sampling of the recent literature in the fields of motor behavior and neurophysiology on which the summary statement in this paragraph is based. It is far from complete.

object. This analogy also points up the ecological adaptation value of a parameter shift mechanism of reinforcement, and may explain the speed with which the shaping process often proceeds.

## 9.5 Explaining Cyclic Behavior Patterns

The FCN operant contingency used in the Mechner, 1958b study (see Sections 3.4 and 8.7) is a quasi-$rO_{FCN}$ under continuous reinforcement, its defining criterion being a certain run length. Figure 2 of that study shows that if the length of a run (the criterial measure) deviates from the mean run length in either direction, then the length of the next run deviates from the mean still more, in the same direction. To the degree that the FCN procedure is an rO, this finding could be regarded as preliminary suggestive evidence for the operation of the parameter shift mechanism.

The parameter shift mechanism also predicts cyclic fluctuations of run lengths, and these are certainly evident in Figure 3 of that study. The lengths of successive reinforced runs should keep shifting away from the mean until a reversal occurs. If we assume that the perpetuation of a parameter shift has a certain probability p which is less than 1.0, then a reversal will soon occur. The average number of consecutive shifts depends on the effective p. At the point of reversal there is a single initial parameter shift in the downward direction, and that is then the direction in which further parameter shifts are perpetuated, until a further reversal occurs. That next reversal, this time from the downward to the upward direction, usually occurs somewhat below the criterion. The average number of consecutive shifts in a given direction depends on p, and would be the same in the upward and downward directions. The average $rO_{FCN}$ run length, which is the mean criterial measure across the cycles, normally tends to fall about 5-10% above the criterion.

The process just described is responsible for the oscillations we see in the lengths of consecutive response runs, and in most other "steady-state" behavior.

## 9.6 The Parameter Shift Mechanism and Behavior Shaping

As was implied above, the parameter shift mechanism may be important during shaping. Shaping procedures usually involve setting progressively more stringent criteria for reinforcement. The dimension along which the criterion is set corresponds to a parameter of the operant. Reinforcers are presented each time the operant's criterial measure has just shifted in the desired direction. When the parameter shift mechanism is operating, the result is a further shift of that parameter in the same direction. This sometimes creates the impression that the subject "understands what is desired," even though the underlying mechanism is quite mechanical. Since the parameter shift mechanism can produce very rapid behavioral changes, it may well be responsible for the speed with which shaping often proceeds.

Reinforcement does not necessarily shift all parameters equally. For example, the force or "overtness" parameter (i.e., degree of muscle engagement)[4] may be impacted more strongly or more frequently than other parameters (Morse, 1966, p. 54). The overtness parameter spans the range from the covert level, where there is no movement at all, to the overt level. When a covert response becomes more forceful, the degree of muscle engagement can reach a level where there is movement, at which point the response is overt. Hefferline & Keenan (1963) showed that when the criterion is a certain thumb muscle potential, operant contingencies can shift the overtness parameter from below to above the threshold for movement.

This would explain how a skilled animal trainer can evoke an operant that may be occurring at a covert level but has not yet occurred overtly. For example, the trainer knows from experience that when the animal fixates an object without yet moving its body, the behavior of moving toward the object may already be occurring at a covert level. A PPRf at that instant tends to impact the overtness parameter of that movement, with the result that an overt movement toward the object may follow.

## 9.7 Questions Regarding the Effects of PPRfs

These are some questions and plausible conjectures that can be investigated with rOs:

(a) What determines *which* operants will be impacted most strongly by a particular PPRf? In Section 9.4 I suggested that it can be the operant whose parameters have just shifted, and that the PPRf's impact is to produce a further shift in the same direction. But a PPRf may shift the parameters of other operants too. For example, it may selectively increase the overtness level of operants that have recently been at high levels of strength, or that have been followed by PPRfs in the past. Thus, a possible variation of the parameter shift mechanism is that a PPRf *produces* a parameter shift in certain operants regardless of whether or not those operants have just had parameter shifts.

(b) How recent is the behavior (or parameter shift) that is repeated when a reinforcer is presented? The experimenter or trainer may have a certain operant and criterion in mind, and presents the reinforcer right after an instance of that operant has occurred. But the reinforcer's impact is not necessarily confined to that operant and that criterial measure (Catania, 1971; 1988). Recency of the targeted behavior may not be the only factor that determines which behavior is impacted. It is plausible that a PPRf can call forth behavior that occurred some time before, including behavior in non-criterial dimensions, because outside the laboratory, operants often produce delayed effects, even when shaping is occurring.

(c) The recency of the behavior (or parameter shift) that is repeated may depend on the recent *density* of PPRfs. Thus, when the PPRf density has recently been high, as in an active

---

[4] We must distinguish between two type of measures: One, which requires multiple instances, is the probability, frequency, or rate of the operant, regardless of its level of force or level of overtness, and the other, which is applicable to single instances, is the operant's level of force or overtness if and when it occurs. It may prove useful to subsume both of these types of measures under the construct of "response strength," but only if it is found that variables that increase one also increase the other.

shaping session, the behavior shifts that are called forth tend to be relatively recent ones, while in situations where PPRfs are sparse, or where there has been no PPRf for a long time, the behavior called forth may tend to be of older vintage.

(d) The effect of a PPRf is not necessarily confined to a single instance of a parameter shift. There may be circumstances in which more than one parameter shift is impacted and perpetuated, i.e., where an entire block of preceding behavior (or parameter shifts) is impacted.

(e) It is possible that the parameter shift mechanism operates only during shaping sessions and not at other times. For example, it has previously been observed that PPRfs, when they occur after a long period without a PPRf, tend to have an arousing or excitatory effect: The subject starts moving faster and more vigorously (Killeen et al., 1978). Once arousal has occurred, the parameter shift mechanism may swing into action, but not until then.

(f) PPRfs have a stronger parameter shift effect on behavior (or parameter shifts) that have received PPRf at least once before, than on parameter shifts that are receiving a PPRf for the first time. In fact, the sensitivity of parameter shifts to PPRfs may increase as a function of the number of times they have previously been followed by PPRfs.

(g) Do PPRfs have different types of effects at different stages of the shaping or automatization process, as the malleability and susceptibility of the behavior undergoes changes or diminishes?

(h) When the parameter shift mechanism is *not* operative, PPRfs may function as $S^D$s only, by selectively evoking behavior that was at high strength in similar situations in the past. As was explained in Section 9.2, every PPRf also functions as part of a compound $S^D$, with the effect of that $S^D$ depending on the subject's earlier history.

(i) PPRfs have a greater impact on parameter shifts that involve overtness increases than overtness decreases. This conjecture is plausible because outside the laboratory, an operant is effective more often when it increases in force, or when a previously covert operant becomes overt, than when it decreases in force or becomes covert.

## 9.8 Effects of Punishment

Punishment may work by *reversing* parameter shifts that were in the direction of greater overtness. It is possible that just as positive reinforcement can *increase* the degree of overtness, punishment can *decrease* it.

Punishment, like positive reinforcement, may affect not only the immediately preceding behavior but an extended block of preceding behavior. That would also explain the well-documented phenomenon of regressive resurgence produced by punishment or stress (Epstein, 1985; Mechner et al., 1992). If punishment has the effect of decreasing the overtness parameter of an extended block of preceding behavior, then the overtness of many individual operants in that block would drop below threshold, and cease to occur overtly or at all. Older behavior would then resurge, because its overtness level would become higher relative to the recently depressed behavior. The automatic result is regressive resurgence.

The punishment literature is replete with statements to the effect that punishment does not alter the *strength* of the punished operant, and suppresses it only temporarily (Keller & Schoenfeld, 1950; Azrin & Holz, 1966). The conjecture that punishment shifts the overtness parameter in the direction of increasing covertness would explain how punishment depresses or suppresses behavior without eliminating it. The conjecture is also plausible from the ecological adaptation standpoint:  The same behavior that is punished in overt form is not punished in covert form ("Think it but don't say it"). Punished operants can occur in covert form and be retained in the behavior repertory for use at a future time when conditions for that behavior may be more favorable.

## 9.9 Implications for the Origins of Response Variability

The normally-observed variability of operant behavior may be due to the continuous action of reinforcements perpetuating the behavior's recent direction of change, thereby producing cyclic fluctuations and sequential effects like those seen in Figures 2 and 3 of Mechner (1958b). We know that during an organism's continuous normal interaction with its environment, large and small reinforcements, in various motivational modalities, constantly impinge on all of its behavior. When the parameter shift mechanism is operative, these reinforcements would generate response variability by continuously shifting the parameters of all ongoing behavior routines in their recent directions of change.[5] The operation of such a mechanism may also explain why the topography of superstitiously conditioned behavior tends to become cyclic, rather than fluctuate randomly (e.g., Skinner, 1948).

Thus, the parameter shift mechanism can explain how reinforcement generates response variants and how it can shape behavior.

## 9.10 A Research Program to Study the Effects of PPRfs

This section outlines the dependent and independent variables of an rO-based research program for studying the types of issues discussed above. The experiments should be done parametrically, because the interactions among the independent variables are likely to be important for the interpretation of the observed effects.

[5] This view of variability follows Sidman's admonition (Sidman, 1960) that variability should not be viewed fatalistically as a manifestation of nature's indeterminacy, and then savored as a comfortingly reliable dependent variable, but should instead be viewed as a scientific challenge, with success measured by the degree to which the observed variations become predictable.

**Dependent variables that should be examined are:**

(a) Comparisons of pre- and post-presentation occurrences of criterial and non-criterial measures. As stated above, the criterial and non-criterial measures represent the parameter settings for each occurrence of the rO. Compare the parameter *shifts* and the parameter *settings* before and after the PPRfs, and determine which (if either) of the two is more strongly perpetuated by the presentation (or reversed if punishment is used). The parameter shift mechanism would produce repetitions of parameter *shifts* rather than parameter *settings*.

(b) Relationships between the dozen-or-so criterial and non-criterial measures that immediately follow the PPRf, and those that occurred earlier in the subject's history, particularly just before and after previous PPRfs. See especially if there is a preponderance, or disproportionate representation, of the non-criterial measures that were most heavily represented in previous blocks of rOs with attention to the possible $S^D$ effects of the stimulus compound that includes the PPRf.

(c) Proximity of a non-criterial measure's position to the end of the rO in which it occurs. A non-criterial measure's sensitivity to the effects of PPRfs may depend on its proximity to the end of the rO (Mechner, et al. 1992).

(d) To investigate the mechanisms of reinforcement at the neurological level, all of the behavioral measures obtained can and should be correlated with concurrently obtained neurological measures.

**Independent variables that should be investigated are**:

*In the Reinforcer Category*

(a) Reinforcer value (e.g., amount of the reinforcer, or motivational level) and valence of the reinforcement. An aversive consequence, like time out or loss of money, can be used instead of positive reinforcement.

(b) Density of PPRfs (a) just prior to the presentation being studied and (b) in previous sessions with which the post-presentation criterial and non-criterial measures are being compared.

(c) Number of times the reinforcer has previously been presented (a) in the subject's history, and (b) in the experiment, under the reinforcement contingency being used.

*In the Pre-Presentation Behavior Category: Type of behavior on which the presentation impinges*

(a) Total number of times the subject has previously emitted that rO.

(b) Behavioral susceptibility, independently measured by one of the methods described in Chapter 8.

(c) Type of rO used (i.e., its classification in terms of the operant contingency).

(d) Time, or number of elapsed rOs, since the last PPRf.

(e) Level of activity (e.g. keystrokes per minute, 1/L, etc.) since the last PPRf, or in the preceding block of rOs.

(f) Reinforcement schedule on which the rO is being, and has previously been, maintained.

*In the Reinforcement Contingency Category*

(a) Reinforcers are presented only when a certain specified shift has just occurred in the selected non-criterial measure. (Note that instituting such a contingency converts the non-criterial measure into a criterial one).

(b) Reinforcers are presented only when a certain selected non-criterial measure has just occurred. (Again, note that instituting such a contingency converts the non-criterial measure into a criterial one).

(c) Repeat *a* and *b* above for criterial measures.

(d) The PPRf is contingent on *two consecutive* shifts in the same direction in two criterial or non-criterial measures. While two successive shifts will be rarer than single shifts, they may be more sensitive to the presentation. It would make sense from the ecological adaptation standpoint for them to be more sensitive, as presentations after double shifts would confirm more selectively and with a firmer basis that the shift is "on the right track."

Note: The criterion for when to present the reinforcer is always based on the shift from a base reading to a comparison reading of criterial or non-criterial measures. The computer makes the determinations by monitoring the shifts on an on-line basis. The base reading can be the average obtained in an immediately-preceding block.

*Procedure used to generate and maintain the baseline behavior*

(a) It is possible to generate a relatively stable stream of rOs by the use of an intermittent reinforcement schedule, like VI or RI, that sustains long stretches of unreinforced rOs. Or a stable stream can be maintained by the presentation of PPRfs in another (weaker?) motivational modality. With human subjects, stable long streams can be sustained by verbal instruction.

(b) Another way is to use continuous reinforcement, with PPRfs that are in the same motivational modality, but where each PPRf consists of an amount of reinforcement that is very small compared to the occasional PPRfs of larger amounts, the latter being the PPRfs that are being studied.