

The effect of varied reinforcement on acquisition and extinction speed

Eduardo Polín and Vicente Pérez
UNED - The National Distance Education University

Abstract

Background: The objective of the present study was to compare both the acquisition speed and the resistance to extinction of a simple discrimination in pigeons, depending on whether the reinforcing consequence included different conditioned reinforcer stimuli or always the same. **Method:** The study was conducted with four experimentally naive female pigeons, which were trained to respond to a “go-no go” procedure in four Skinner boxes with coupled touchscreens. The subjects were divided into two conditions with two subjects each. In the Experimental condition, the reinforcement of correct trials was performed by the presentation of one of four previously trained conditioned reinforcers followed by food, while in the Control condition the conditioned reinforcer stimulus was always the same. After acquiring the discrimination, all the subjects were exposed to the extinction phase. **Results:** The subjects of the Experimental condition needed about half of the sessions that the subjects of the Control condition needed to acquire the discrimination. In addition, subjects of the Experimental condition continued responding for more sessions than Control subjects in the extinction phase, although there were no differences in the resistance to extinction. **Conclusions:** Acquisition speed is greater when a variety of conditioned reinforcers is applied.

Keywords: Variety of conditioned reinforcers, simple discrimination, acquisition, extinction, pigeons.

Resumen

Efecto del reforzamiento variado en la velocidad de adquisición y extinción. **Antecedentes:** el presente estudio tenía como objetivo comparar tanto la velocidad de adquisición como la resistencia a la extinción de una discriminación simple en palomas en función de si la consecuencia reforzante comprendía distintos estímulos reforzadores condicionados o siempre el mismo. **Método:** el estudio se realizó con cuatro palomas hembra experimentalmente ingenuas entrenadas para responder a un procedimiento “go-no go” en cuatro cajas de Skinner con pantallas táctiles acopladas. Los sujetos fueron divididos en dos condiciones con dos sujetos cada una. En la condición Experimental el reforzamiento de los ensayos correctos se realizó mediante la presentación de alguno de los cuatro reforzadores condicionados previamente seguido de comida, mientras que en la condición Control el reforzador condicionado era siempre el mismo. Tras adquirir la discriminación, todos los sujetos pasaron a la fase de extinción. **Resultados:** los sujetos de la condición Experimental necesitaron aproximadamente la mitad de sesiones para adquirir la discriminación que los de la condición Control. Además, los sujetos de la condición Experimental continuaron respondiendo durante más sesiones que los de la condición Control durante la fase de extinción, aunque ambos presentaron la misma resistencia. **Conclusiones:** la velocidad de adquisición es mayor cuando se aplica una variedad de reforzadores condicionados.

Palabras clave: variedad de reforzadores condicionados, discriminación simple, adquisición, extinción, palomas.

In simple discrimination training, the subject is exposed to two or more stimuli whose appearance correlates differently with the reinforcement of a certain behavior (De Rose, McIlvane, Dube, Galpin, & Stoddard, 1988, for example). Thus, if the subject emits the target behavior in the presence of the Discriminative Stimulus (DS or S+) the consequence will be reinforcing, while if this behavior is emitted in the presence of the Delta Stimulus (Δ S or S-) it will be punished or extinguished.

Overall, it could be said that the discrimination between two events (measured by the emission or not of the target behavior) is acquired faster the more different the elements involved in each

of the terms which compose the contingencies are from each other: the Discriminative Stimuli (White, Pipe, & McLean, 1985), the responses (Cohen, Looney, Brady, & Aucella, 1976) and the consequences (Fedorchak & Bolles, 1987; Peterson, Wheeler, & Armstrong, 1978).

The present study focuses on some issues concerning the consequences. On this topic there is an important variable which has received remarkable attention within the study of discriminations learning: the use of conditioned reinforcers (Williams, 1994, for a theoretical review).

We consider an initially neutral stimulus as a conditioned reinforcer when it acquires its function by being paired with another reinforcing event (Pellón, Miguéns, Orgaz, Ortega, & Pérez, 2014). We can find examples of conditioned reinforcers in attention, congratulations, money or points.

The effect of including a conditioned reinforcer with a primary one affects the maintenance of the response (Hyde, 1976) and also its acquisition speed (Williams & Dunn, 1991). Hursh (1977) also

observed, in monkeys, the increase of a simple discrimination acquisition speed, adding a conditioned reinforcer before the appearance of the primary one. But, because of the features of his procedure, in which another response must be emitted following the reinforcer appearance, the interpretation of the results was based on the discriminative function of these stimuli. Other studies (Caggiula et al., 2002; Chaudhri et al., 2005; Donny et al., 2003), using nicotine as a primary reinforcer, have found that the acquisition speed increases when a conditioned reinforcer is introduced.

The objective of the present study, however, was to check if learning is also faster as a function of the variety of conditioned reinforcers involved in the process, and not just by the mere inclusion of a conditioned reinforcer before the primary one.

Several authors have shown interest in studying the different behavioral effects of varied versus constant reinforcement. In this regard, there is some evidence pointing to the fact that learning could be favored by applying a variety of reinforcers.

For example, Egel (1980) reported, in a study done with autistic children, that applying a variety of primary reinforcers leads to a larger number of responses in comparison with constant reinforcement. The same author found that varied (versus constant) primary reinforcement is also associated with higher levels of correct responding (Egel, 1981) and the results were discussed in terms of motivation (greater satiation with regard to the constant reinforcer).

An experiment performed by Bowman, Piazza, Fisher, Hagopian, and Kogan (1997) with 7 children who had been diagnosed with moderate to profound mental retardation, shows that preference for varied (versus constant) reinforcers could be established even when the quality of the varied reinforcers is lower than the constant reinforcer (although 2 participants preferred the constant, higher quality reinforcer, and one showed no systematic preference for either option).

In a more recent study, Milo, Mace, and Nevin (2010) assessed the preference, the response rates and the resistance to change when comparing varied and constant reinforcement. Their results suggest, in line with previous literature, that varied reinforcement maintains higher response rates than constant reinforcement and that it is preferred over constant reinforcement. This time, the results concerning preference for varied reinforcement (with 4 out of 4 participants preferring this option) were more reliable than those reported by Bowman et al. (1997). Moreover, they found that varied reinforcement also leads to a greater resistance to change in a distraction task.

There is also some evidence in non-human animal basic research suggesting that a variety of reinforcers maintains higher rates of response than constant reinforcers (Steinman, 1968a; 1968b), although there are other authors who claim that this varied reinforcement effect depends on specific experimental manipulations as well as on the possible interaction between the particular reinforcers used (Roca, Milo, & Lattal, 2011).

The present study was designed to check if learning is not only favored in terms of response rate or preference, but also in terms of acquisition speed and resistance to extinction, by applying a variety of consequent events following the emission of an operant response. Furthermore, in all the cited studies about varied versus constant reinforcement only primary reinforcers (different edible stimuli) were used, therefore, in the present study we used visual conditioned reinforcers, pairing all of them with the same kind of food, so that a hypothetical differential effect due to the varied

(or constant) reinforcement could not be explained in terms of satiation with regard to the primary reinforcer.

Method

Participants

Four rock pigeons (*Columba livia*) were kept at approximately 80% of their Free-Feeding Weight (FFW). Water and pigeon grit were always available in their home cages. The room was maintained at a stable temperature and humidity and a 12:12 hour light-dark cycle was employed during the whole experiment (the lights turned on at 10 am).

Instruments

Four modified operant chambers for bird conditioning with a tactile PC screen attached on the left side were used. Each chamber was 43.5 cm high, 64 cm long and 45 cm wide. The front of the chamber was equipped with a feeder providing a mix of grain and with three standard keys (left, central, right). In the middle of the back side, a 35 watt white light provided illumination. The left panel of the chamber was removed, and a touch screen monitor was installed instead (ELO Touchsystems MODEL ETL 121-C-75WB-1). The monitor displayed a resolution of 800 per 600 pixels, 60Hz, SVGA (16.2 million colours).

The modified operant chambers were enclosed in sound-attenuating hulls. A ventilation fan installed in each hull produced a white noise, masking extraneous sounds.

Each touch screen was connected to an IBM compatible PC. A tailor made program (DV) controlled the presentation of stimuli and recorded the responses. Each PC was connected to a MED R/M interface cabinet SG - 6001C SN controlled by an IBM compatible computer. The software used to run the experiments was MedPC 2.0 for Windows.

The stimuli appeared through the keys and through the screen. The side keys (left and right) could be illuminated either green (Discriminative Stimulus, DS or S+) or red (Delta Stimulus, ΔS or S-). Another five stimuli were presented through the touch screen monitor (only one at the same time). They consisted of images of polygonal shapes of different colours on a black background.

Procedure





















Before the beginning of the experiment, all the pigeons were auto-shaped until a white central key-pecking response was established at a consistent level.

Phase 1: Conditioned reinforcers acquisition. This phase, which was the same for all the subjects, consisted of the appearance of the five stimuli (only one per trial) through the touch screen. Stimuli CS1, CS2, CS3 and CS4 were paired with food presentation. Each of them maintained a different contingency with the appearance of food (mean of 0.9), counterbalanced for each subject (see Table 1). Stimulus CS5 was paired with no food appearance (-0.9). Sessions consisted of 65 trials, each stimulus being presented an average of 13 times per session.

The finishing criteria for this phase were the appearance (and maintenance) of a stable level of touchscreen-pecking conditioned response (CR) in the presence of CS1, CS2, CS3 and CS4 stimuli and the lack of this response in the presence of CS5.

Table 1

Stimuli used for each subject in phase 1 as a function of their contingency relationship with the food presentation. Although the stimuli appear in grey scale, they were originally coloured

Subject	CS1 (0.9)	CS2 (0.9)	CS3 (0.8)	CS4 (1)	CS5 (-0.9)
1					
2					
3					
4					

Phase 2: Simple discrimination acquisition. The subjects were divided into two conditions: Experimental (variety) and control (constant). A Fixed Ratio (FR) 3 schedule was designed, using a “go/no go” successive simple discrimination procedure in which the DS (green light) and ΔS (red light) appeared an average of 50% of the trials each, counterbalanced by position (left and right). For both groups, sessions consisted of 40 intermixed trials, therefore, DS and ΔS were presented an average of 20 times each per session with a 45 second inter-trial interval (ITI).

Both DS and ΔS remained activated until the subjects responded or until 10 seconds passed. We chose to require white central key-pecking rather than pecking directly to the DS in order to isolate operants and conditioned responses, so, when the DS or the ΔS appeared, the central key was white illuminated until the response was given or until 10 seconds passed.

Trials in which subjects responded pecking to the white central key in the presence of the DS, as well as trials in which they did not respond in the presence of the ΔS, were considered correct. Those trials in which subjects responded pecking to the white central key in the presence of the ΔS, and those trials in which they did not respond in the presence of the DS, were considered failures.

For the Experimental condition subjects, correct trials were followed by the random appearance of one of the four previously conditioned reinforcers (CS1, CS2, CS3 or CS4). These stimuli remained on the screen eight seconds and were always followed by five seconds of access to the feeder. Failure trials were punished by the appearance of CS5 for eight seconds and no food was presented. For the Control condition subjects, the same procedure was used, but, correct trials were always followed by the appearance of the same conditioned reinforcer (CS1) and food.

The number of sessions needed to reach different acquisition criteria (75, 85, 90, and 95% of correct trials for a whole session) was measured as a dependent variable.

All the pigeons had a total of 61 sessions, regardless of the acquisition speed shown by each one.

Phase 3: Simple discrimination extinction. The contingencies were the same as in the previous phase for both groups, but, in this time regardless of which stimulus was present and the behavior of the subjects, no conditioned reinforcers and no food appeared.

All the subjects had a total of 18 sessions and each session consisted of 40 intermixed trials, with 45 seconds of ITI. The

number of sessions needed to return to the random level (50% of correct and failure trials for a whole session) was measured as a dependent variable.

Data analysis

Given the nature of the data, as well as the number of subjects, the analysis strongly relied on graphic representation and visual analysis.

Results

Simple discrimination acquisition phase

All the subjects acquired the simple discrimination during the 61 sessions. Differences between the two conditions can be observed when comparing the number of sessions needed to reach the different acquisition criteria (75%, 85%, 90%, and 95% of correct trials for a whole session). Figure 1 shows these results.

Figure 2 depicts the discrimination acquisition mean curve for each condition and individually for each subject.

Simple discrimination extinction phase

The simple discrimination was extinguished for all the subjects during the 18 sessions. No differences were found when comparing the number of sessions needed to return to the random level of response. Figure 3 depicts these results.

However, there is a slight difference between subjects in the reduction percentage of the key-pecking response during this phase (with regard to the number of responses given in their respective first extinction sessions). Figure 4 presents these results.

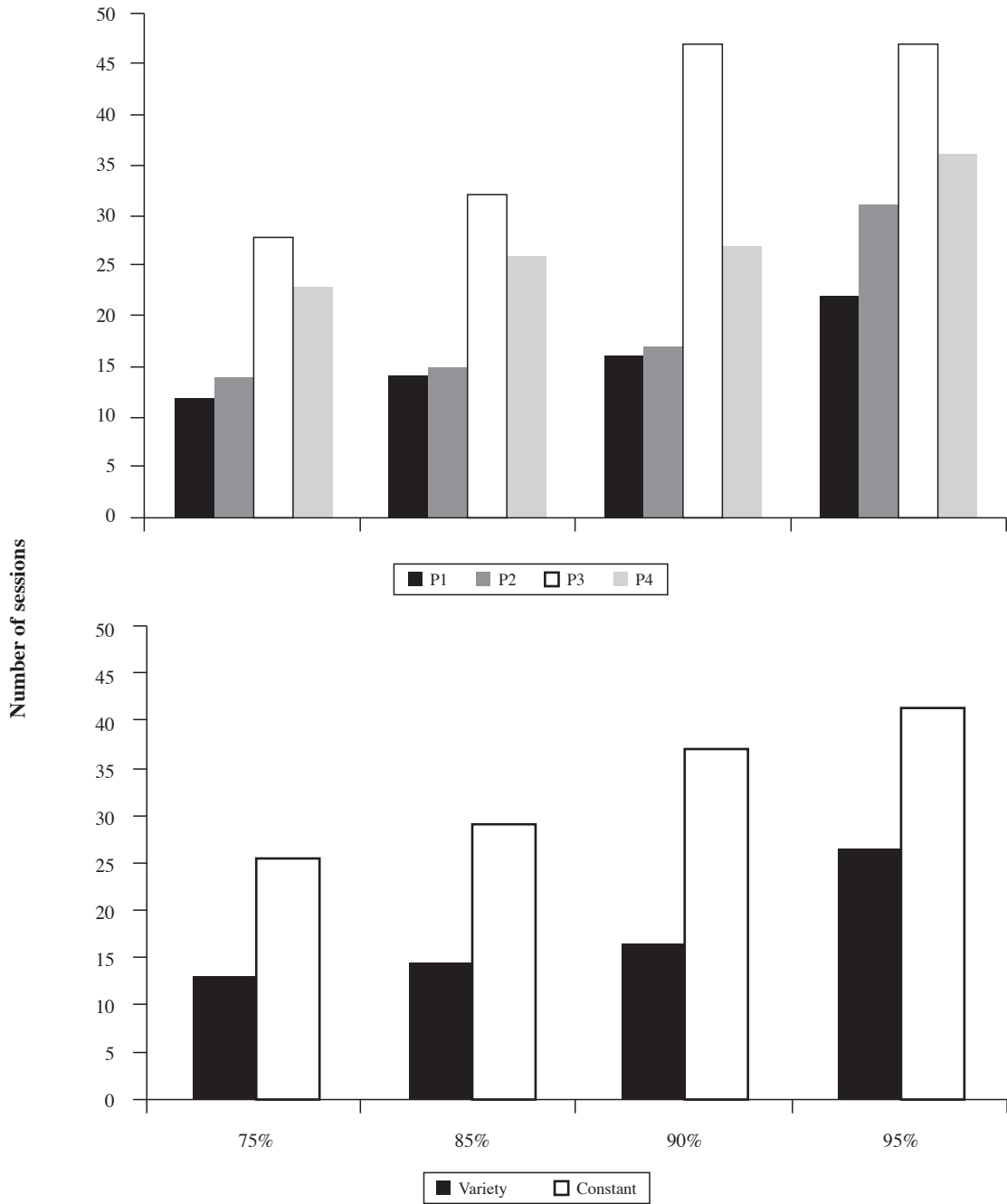
Specifically, the subjects of the Experimental condition (variety) responded more than the subjects of the Control condition (constant) with regard to their respective baselines, although the accuracy of the discrimination was similar for all the subjects.

Discussion

We can highlight one main conclusion from the acquisition phase results: The data show an acceleration of learning in the experimental condition compared with the control; that is, subjects of the Experimental condition (variety) required fewer sessions than subjects of the Control condition (constant) to reach the different acquisition criteria. These findings support the initial hypothesis, which predicted that learning would be favored, in terms of acquisition speed, by the inclusion of a variety of conditioned reinforcers (over the use of only one) before the appearance of the food.

This finding is consistent with previous studies (Bowman et al., 1997; Egel, 1980, 1981; Milo et al., 2010; Steinman, 1968a, 1968b) and, furthermore, provides evidence that a varied reinforcement effect can be observed even when there are no motivational differences between subjects. All the pigeons were kept at the 80% of their FFW and were reinforced with the same kind of food. In addition, the conditioned reinforcers were also paired with the same kind of food, so, it does not seem reasonable to consider that the differential effect found could be due to satiety issues.

A possible explanation for these results starts out from considering that the DS, in turn, can acquire conditioned



Acquisition criteria (% of correct trials)

Figure 1. Number of sessions needed to reach the different acquisition criteria. Individual subject data and mean values as a function of the condition. Pigeons 1 and 2 belong to the experimental condition (variety). Pigeons 3 and 4 belong to the control condition (constant)

stimulus functions (that is, could function as a conditioned reinforcer), increasing the response rate in its presence (Williams & Dunn, 1991). This possibility is supported by studies like Dinsmoor (1950), for example, who managed to reinforce a behavior by the contingent presentation of a stimulus which had previously functioned as a discriminative stimulus for another behavior.

Following this hypothesis, it could be argued that the appearance of a greater number of signals turns the DS into a much more relevant event for the subject. Therefore, it would be more adaptive that the behavior fell under its control as quickly as possible. This would explain the acquisition speed differences found.

Along the same lines, this phenomenon could be related to the so-called generalized conditioned reinforcer. This special kind of

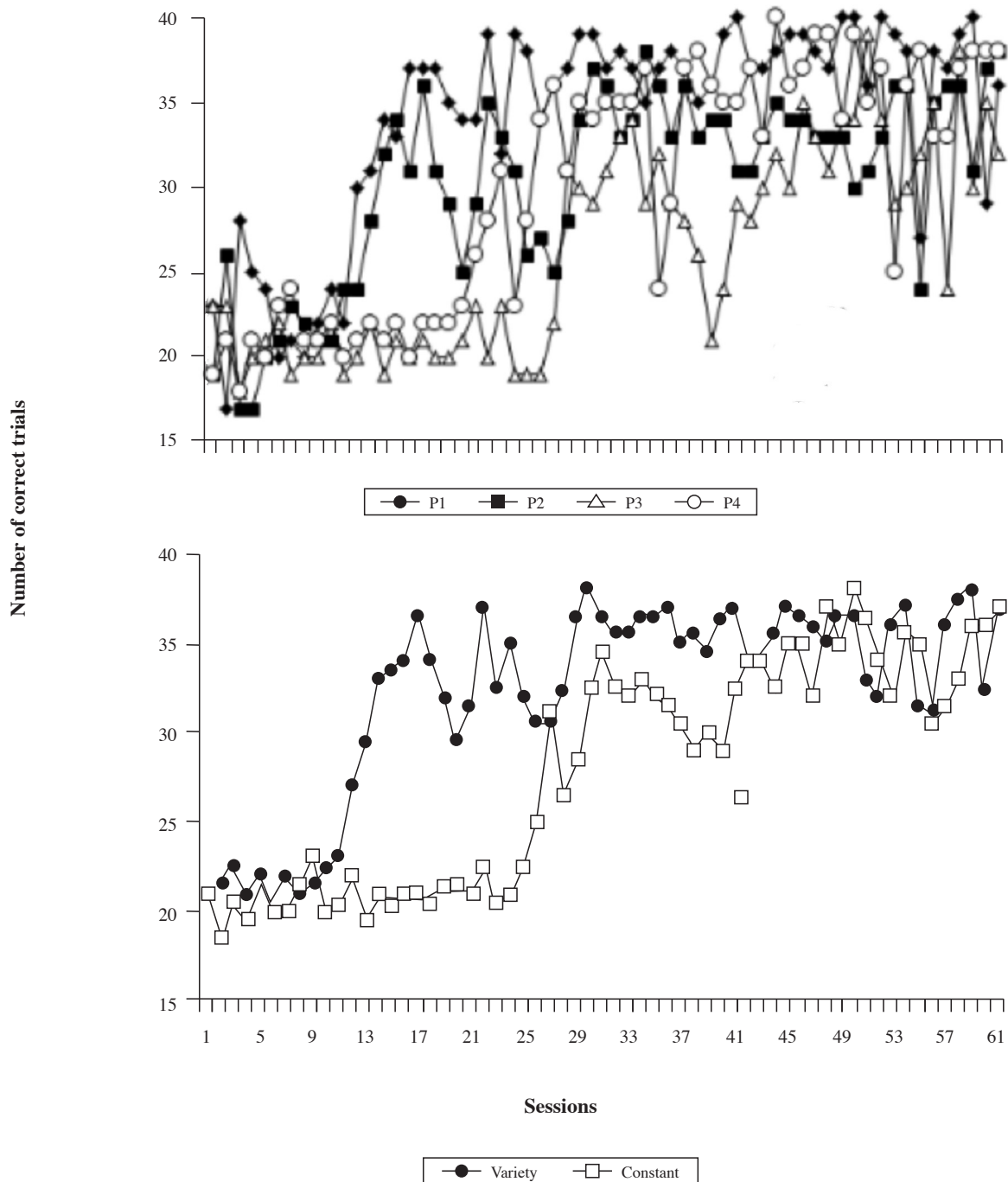


Figure 2. Number of correct trials along the sessions of the simple discrimination acquisition phase. Individual subject data and mean curves as a function of the condition

conditioned reinforcer acquires its function by being paired with a wide variety of events which have already functioned as reinforcers (Lamarre & Holland, 1985; Skinner, 1953, for example). The main feature of this kind of stimuli is that they are not dependent on the deprivation or satiation level of any particular primary reinforcer with which they have been paired. We can see examples of these kinds of events in most social reinforcers, such as attention, affect, approval, money, etc.

It could be considered, making an analogy, that those DS which signal with a high probability that a behavior will be reinforced with a wide variety of reinforcers (compared to a single one) also acquire characteristic properties. We can find an example that meets this description in the sound of our names. This event correlates in many different ways with the reinforcement of the behavior of orientation towards the source of the sound: avoiding being hit by a car, finding a loved one, taking your turn

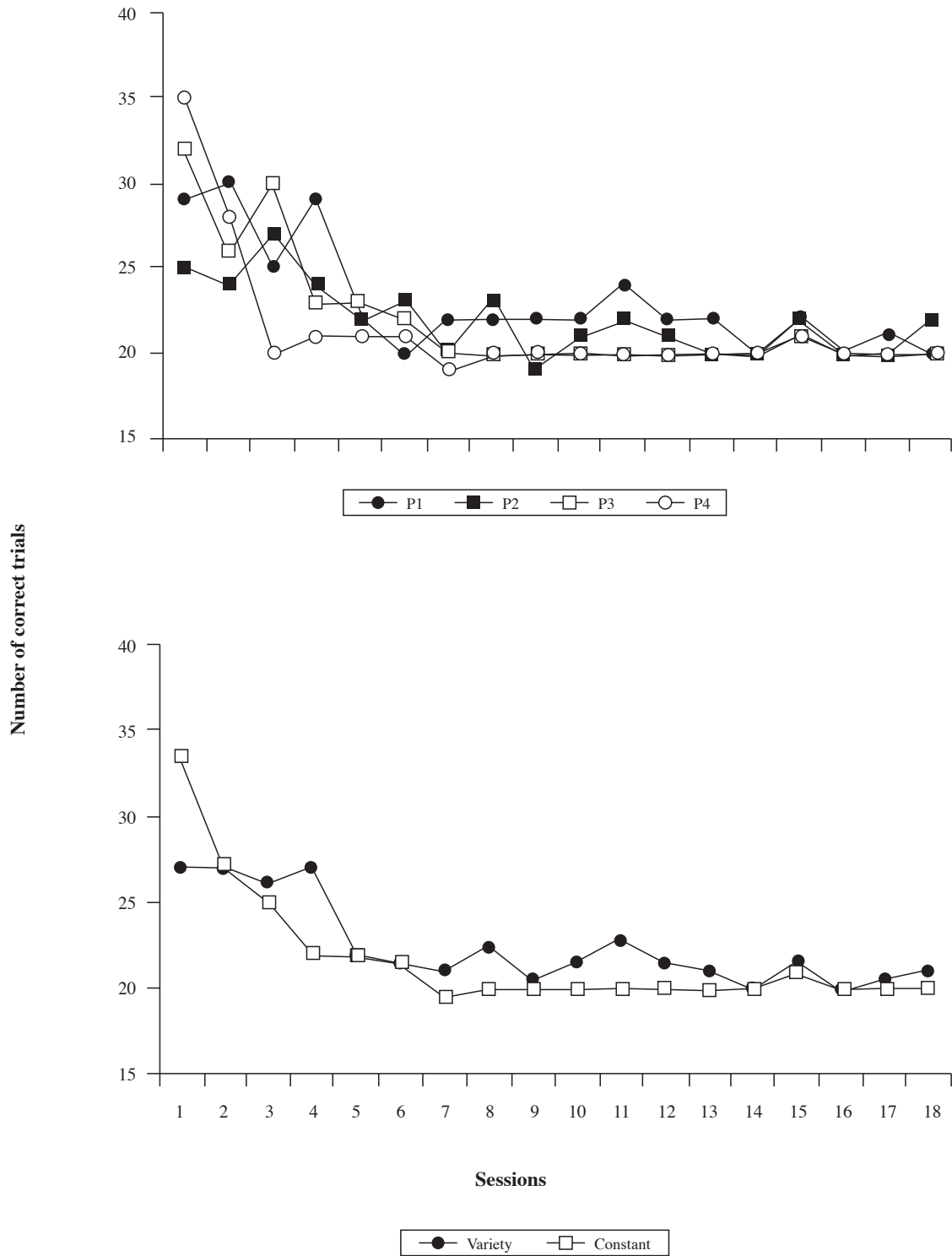


Figure 3. Number of correct trials along the sessions of the simple discrimination extinction phase. Individual subject data and mean curves as a function of the condition

in a store or in the doctor’s office etc. Our orientation response following the sound of our name is a behavior that, like the generalized conditioned reinforcer effect, is usually emitted with a high probability regardless of the subject deprivation level and, presumably, is acquired with certain speed.

The results of the extinction phase are not so clear: On one hand, there are no differences between the subjects of the two conditions in the number of sessions required to return to a random level of responding. The starting hypothesis predicted that the discriminative control would take longer to be extinguished in

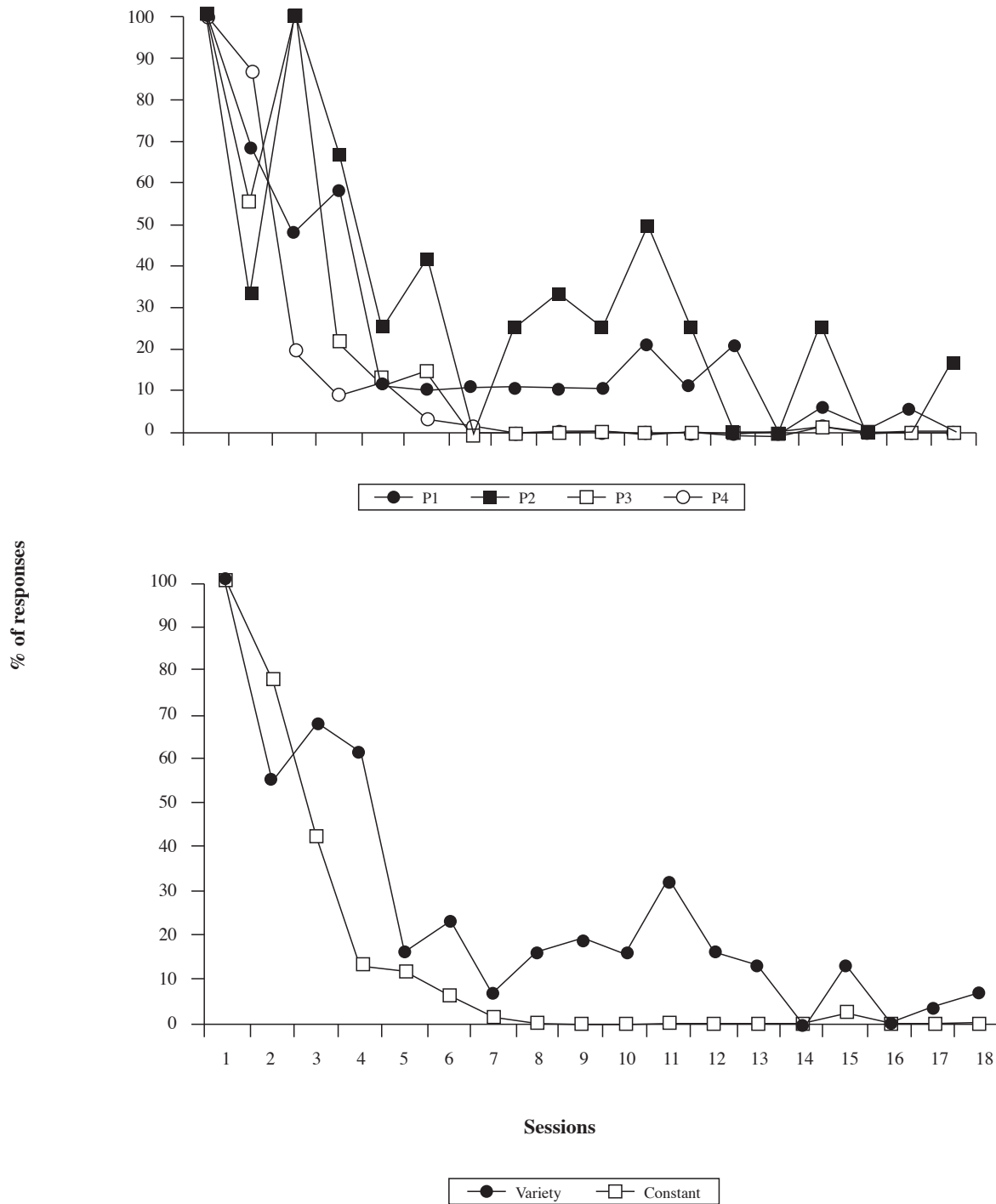


Figure 4. Reduction of the key-pecking response along the extinction phase for each subject and mean values as a function of the condition. Y-axis shows the response percentage, being 100% the whole number of responses given by each subject in their respective first session. The number of responses given by each subject in the first session (100%) was: Pigeon 1: 19; Pigeon 2: 12; Pigeon 3: 27; Pigeon 4: 67

the Experimental condition than in the Control, but the data do not support this claim. This was probably due to the fact that the independent variable introduced in the acquisition phase (varied or constant reinforcement) was not present during this one, that is to say, in the extinction phase the contingencies were exactly the same in both conditions (neither conditioned nor primary

reinforcers appeared). Further research is needed in order to check if the same result would be obtained when continuing presenting the conditioned (but not the primary) reinforcers in the extinction phase as well.

On the other hand, there is a slight difference between the subjects of the two conditions in the reduction percentage of the

responses given in the extinction phase with regard to the baselines. More specifically, the total amount of responses per session given by the subjects of the Experimental condition decreased slower than those given by the subjects of the Control condition, although, as said before, the accuracy of the discrimination was not affected. These data are consistent with the findings reported by Milo et al. (2010), which suggest that varied reinforcement maintains higher

response rates and also leads to a greater resistance to distraction than constant reinforcement.

In summary, although it is necessary to keep on looking into the features and sturdiness of this phenomenon more deeply, the confirmation that applying a variety of reinforcers in training improves learning speed may have important educational and training implications.

References

- Bowman L. G., Piazza C. C., Fisher W. W., Hagopian L. P., & Kogan J. S. (1997). Assessment of preference for varied versus single reinforcers. *Journal of Applied Behavior Analysis, 30*, 451-458.
- Caggiula, A. R., Donny, E. C., White, A. R., Chaudhri, N., Booth, S., Gharib, M. A., ..., Sved, A. F. (2002). Environmental stimuli promote the acquisition of nicotine self-administration in rats. *Psychopharmacology, 163*, 230-237.
- Chaudhri, N., Caggiula, A. R., Donny, E. C., Booth, S., Gharib, M. A., Craven, L. A., ..., Perkins, K. A. (2005). Sex differences in the contribution of nicotine and nonpharmacological stimuli to nicotine self-administration in rats. *Psychopharmacology, 80*, 258-266.
- Cohen, L. R., Looney, T. A., Brady, J. H., & Aucella, A. F. (1976). Differential sample response schedules in the acquisition of conditional discriminations by pigeons. *Journal of the Experimental Analysis of Behavior, 26*, 301-314.
- De Rose, J. C., McIlvane, W. J., Dube, W. V., Galpin, V. C., & Stoddard, L. T. (1988). Emergent simple discrimination established by indirect relation to differential consequences. *Journal of the Experimental Analysis of Behavior, 50*, 1-20.
- Dinsmoor, J. A. (1950). A quantitative comparison of the discriminative and reinforcing functions of a stimulus. *Journal of the Experimental Psychology, 40*, 458-472.
- Donny, E. C., Chaudhri, N., Caggiula A. R., Evans-Martin, F. F., Booth, S., Gharib, M. A., ..., & Sved, A. F. (2003). Operant responding for a visual reinforcer in rats is enhanced by non-contingent nicotine: Implications for nicotine self-administration and reinforcement. *Psychopharmacology, 169*, 68-76.
- Egel, A. L. (1980). The effects of constant vs. varied reinforcer presentation on responding by autistic children. *Journal of Experimental Child Psychology, 30*, 455-463.
- Egel, A. L. (1981). Reinforcer variation: Implications for motivating developmentally disabled children. *Journal of Applied Behavior Analysis, 14*, 345-350.
- Fedorchak, P. M., & Bolles, R. C. (1987). Hunger enhances the expression of calorie-but not taste-mediated conditioned flavor preferences. *Journal of Experimental Psychology: Animal Behavior Processes, 13*, 73-79.
- Hursh, S. R. (1977). The conditioned reinforcement of repeated acquisition. *Journal of the Experimental Analysis of Behavior, 27*, 315-326.
- Hyde, T. S. (1976). The effect of Pavlovian stimuli on the acquisition of a new response. *Learning and Motivation, 7*, 223-239.
- Lamarre, J., & Holland, J. G. (1985). The functional independence of mands and tacts. *Journal of the Experimental Analysis of Behavior, 43*, 5-19.
- Milo, J. S., Mace, F. C., & Nevin, J. A. (2010). The effects of constant versus varied reinforcers on preference and resistance to change. *Journal of the Experimental Analysis of Behavior, 93*, 385-394.
- Pellón, R., Miguéns, M., Orgaz, C., Ortega, N., & Pérez, V. (2014). *Psicología del Aprendizaje [Psychology of Learning]*. Madrid: UNED.
- Peterson, G. B., Wheeler, R. L., & Armstrong, G. D. (1978). Expectancies as mediators in the differential-reward conditional discrimination performance of pigeons. *Animal Learning & Behavior, 6*, 279-285.
- Roca, A., Milo, J. S., & Lattal, K. A. (2011). Efectos del reforzamiento cualitativamente variado sobre la tasa de respuesta en ratas [Effects of qualitatively varied reinforcement on response rate in rats]. *Acta Comportamentalia, 19*, 3-18.
- Skinner, B. F. (1953). *Science and human behavior*. New York: Macmillan.
- Steinman W. M. (1968a). Response rate and varied reinforcement: Reinforcers of similar strengths. *Psychonomic Science, 10*, 35.
- Steinman W. M. (1968b). Response rate and varied reinforcement: Reinforcers of different strengths. *Psychonomic Science, 10*, 36.
- White, K. G., Pipe, M. E., & McLean, A. P. (1985). A note on the measurement of stimulus discriminability in conditional discriminations. *Bulletin of the Psychonomic Society, 23*, 153-155.
- Williams, B. A. (1994). Conditioned reinforcement: Experimental and theoretical issues. *The Behavior Analyst, 17*, 261-285.
- Williams, B. A., & Dunn, R. (1991). Substitutability between conditioned and primary reinforcers in discrimination acquisition. *Journal of the Experimental Analysis of Behavior, 55*, 21-35.