Revista Mexicana de Análisis de la Conducta • Mexican Journal of Behavior Analysis
2019 | Núm. 2 • diciembre | No. 2 • December | Vol. 45, 149 - 172
http://doi.org/10.5514/rmac.v45.i2.75561

B. F. SKINNER'S EVOLVING VIEWS OF PUNISHMENT: I. 1930-1940

LA EVOLUCIÓN DEL PUNTO DE VISTA DE B. F. SKINNER SOBRE EL CASTIGO: I. 1930-1940

Bruna Colombo dos Santos
Universidade Estadual de Feira de Santana–UEFS/Pontifícia Universidade Católica de São Paulo – PUC

Marcus Bentes de Carvalho Neto
Universidade Federal do Pará–UFPA

Abstract

Punishment is a controversial topic. In the theoretical field, there are two definitions of punishment that correspond to two theories: one that considers punishment as asymmetric to reinforcement and the other that considers it symmetric. One of authors that defended an asymmetric view was B. F. Skinner. Citations of Skinner's

Bruna Colombo dos Santos (Universidade Estadual de Feira de Santana–UEFS/Pontifícia Universidade Católica de São Paulo–PUC) Marcus Bentes de Carvalho Neto (Universidade Federal do Pará–UFPA)

The authors thank the Brazilian Federal Agency for the Support and Evaluation of Graduate Education (CAPES) for financing this research through a doctoral scholarship and sandwich doctorate granted to the first author. The authors also thank professor Dr. Carlos Souza (UFPA) who improved the quality of this paper and the B. F. Skinner Foundation for support for the first author.

Address: Rua Augusto Corrêa, 01, Campos Universitário do Guamá, Belém, Pará, Brazil, CEP: 66075110. Phone: (91) 3201-8542/8476. Email: brucolombodossantos@gmail.com

position on punishment most often rely on what he described in *Science and Human Behavior*. The objective of this review was to present the historical development of the concept of punishment in B. F. Skinner's work, in the early years of his career, from the 1930s. We consider the definition, explanatory mechanisms, concepts related to punishment and the notion of symmetry and asymmetry. The term used in the 1930s to refer to punishment was negative reinforcement/conditioning. Skinner talked about punishment for the first time in 1935, considering it a process that decreased operant strength. In 1938, he began questioning this punishment effect, culminating in a change in definition in the latter year. The possible reasons for this change were the development of the concept of reserve and Konorski and Miller's (1937) criticisms.

Key words: punishment, behavior analysis, B. F. Skinner, history

Resumen

El castigo es un tema controvertido. En el campo teórico, hay dos definiciones de castigo que corresponden a dos teorías: una que considera el castigo como asimétrico al refuerzo y la otra que lo considera simétrico. Uno de los autores que defendió una visión asimétrica fue B. F. Skinner. Las citas de la posición de Skinner sobre el castigo a menudo dependen de lo que describió en Science and Human Behavior. El objetivo de esta revisión fue presentar el desarrollo histórico del concepto de castigo en el trabajo de B. F. Skinner, en los primeros años de su carrera, desde la década de 1930. Consideramos la definición, los mecanismos explicativos, los conceptos relacionados con el castigo y la noción de simetría y asimetría. El término utilizado en la década de 1930 para referirse al castigo era refuerzo/condicionamiento negativo. Skinner habló sobre el castigo, por primera vez, en 1935, considerándolo un proceso que disminuyó la fuerza operante. En 1938, Skinner comenzó a cuestionar este efecto de castigo. Observamos un cambio en la definición que ocurrió entre 1935 y 1938. Las posibles razones de este cambio fueron: el desarrollo del concepto de reserva y las críticas de Konorski y Miller (1937).

Palabras clave: castigo, análisis de la conducta, B. F. Skinner, historia

Punishment is a controversial subject in the theoretical, experimental, and applied components of behavior analysis. There are at least two types of theoretical definitions of punishment. One is classified as procedural (Hineline, 1984; Skiba

& Deno, 1991), where punishment is defined strictly on the basis of operations — the addition of a negative reinforcer and the removal of a positive reinforcer (e.g., Skinner, 1953/2005). The other type is classified as functional, where punishment is defined based on behavioral change (i.e., decrease in probability) produced by the procedure of adding negative reinforcers and removing positive ones contingent on responding (e.g., Azrin & Holz, 1966; Catania, 1999). Such definitions are related to two distinct theories of punishment: one that considers punishment asymmetrical in relation to reinforcement and the other that considers punishment symmetrical to reinforcement (Holth, 2005).

In the asymmetrical view, punishment is not considered a behavioral process in the first place. Punishment is considered only a procedure and its effects are explained by other behavioral processes (i.e., negative reinforcement). In the symmetrical view, punishment, as well as reinforcement, is considered an independent behavioral process the effect of which is symmetrically opposed to reinforcement. That is, if reinforcement increases the frequency of a response class, punishment decreases that frequency. The effect of punishment on behavior does not need to be explained by another behavioral process, punishment itself is sufficient.

Although authors such as Michael (1975), Skiba and Deno (1991), Lerman and Vorndran (2002) and Holth (2005) have suggested that the symmetrical position is most often referred to by behavior analysts, this has not eliminated the asymmetrical position, much less the debates about it. Thus, both theories still coexist and continue to be debated by behavior analysts (e.g., Spradlin, 2002; Hoth, 2005; Gongora, Mayer & Mota, 2009; Mayer & Gongora, 2011; Carvalho Neto & Mayer, 2011; Hineline & Rozales-Ruiz, 2013). In regard to the asymmetrical approach, these investigations have considered more specifically the position assumed by Skinner (1953/2005) to evaluate his ideas regarding punishment, critically demonstrating its terminological and conceptual characteristics, and comparing it to the symmetrical approach (Hoth, 2005; Gongora, Mayer & Mota, 2009; Mayer & Gongora, 2011; Carvalho Neto & Mayer, 2011).

It is worth mentioning that Skinner (1953/2005) is not the only author who has presented an asymmetrical position. Authors like Thorndike (1931), in the weak law effect, Estes (1944/1968), Dinsmoor (1954; 1955; 1977; 1998), Solomon (1964) and Sidman (1989/1995) also have promulgated approaches closely aligned with this view. These approaches, however, are not identical to Skinner's.

Nonetheless, commentators predominantly analyze Skinner's position. This seems justified by the central role that Skinner has had in creating and consolidating

behavior analysis; and because in the literature discussing punishment, regarding theoretical aspects as much as ethical ones, Skinner is frequently referred to (Hineline, 1984; Griffin, Paisey, Stark & Emerson, 1988; Skiba & Deno, 1991; Mayer & Gongora, 2011; Martins, Carvalho Neto & Mayer, 2013). Hence, Skinner's views remain influential and worthy of further review.

In investigating the secondary literature analyzing Skinner's view on punishment (Carvalho Neto & Mayer, 2011; Gongora, Mayer & Mota, 2009; Hoth, 2005; Mayer & Gongora, 2011), the studies mainly use his 1953 analysis as support in regard to describing and analyzing central aspects of this concept. *Science and Human Behavior* is one of the most important texts in behavior analysis and indeed it does present a detailed analysis of punishment. However, examining one author's position about a certain concept largely using only one piece of his work may lead to conceptions out of the context of the complete theoretical system. Thus, the role that other concepts may have had in the creation of the concept being examined, as well as the historical nuances related to its construction and possible changes over time, end up being overlooked.

Viera Pinto (1979) supported the notion that the investigation of scientific ideas, whether in general or philosophical, is only possible through their historical development. In his words, "the content of every concept is the process of its conceptualization" (p. 91). Beyond the historical aspect, Viera Pinto affirmed that no concept can be understood singly, without comprehending other concepts that were likely to be present in the formulation of the concept studied.

Therefore, the objective of this review is to present the formulation of Skinner's concept of punishment from the beginning of his journey as a behavioral scientist in the 1930s. The historical development of the concept between 1931 and 1940 are examined. His definition of punishment is considered along with how he explained the behavioral suppression produced by punishment, concepts related to his definition and to the explanation of punishment, and issues of symmetry and asymmetry. These issues were not randomly selected, and the issues that have arisen around them are there because they are still present in current debates about punishment. Thus, to question them historically makes sense because gaps and controversies still exist (Araujo, 2016).

Historical research is justified, among other reasons, for filling gaps in a discipline, helping to solve ongoing dilemmas and aiding in the comprehension of how the discipline has become what it is (Morris, Todd, Midgley, Schneider, & Johnson, 1990; Rampolla, 2015). Hence, it is hoped that through the initial formulation of

the concept of punishment, it will be possible to identify if there were changes in Skinner's ideas, which elements led him to explain punishment as he did, and how it helps in understanding his formulation of punishment in subsequent decades.

The initial years: a brief contextualization of the 1930s

Skinner's behavioral system has changed since its inception. If we are to understand punishment development, we need to understand how his system changed over time. This seems to be important, because the reader will see a lot of "reflex terminology" in the 1930's being used to describe which we currently call operant processes. This was true at that time, because Skinner's system was conceptualized in terms of reflex laws. In this section we describe some of the general characteristics of Skinner's system in the 1930s to define and clarify his terminology that appears in subsequent sections of the current review. Skinner's (1931/1999) conceptualization of the reflex is the initial milestone of his system. He examined the concept of the reflex and its historical formulation and proposed an alternative definition compatible with his objectives of establishing an independent science of behavior. The reflex was defined as "an observed correlation of two events, a stimulus and a response" (Skinner 1931, p. 494). The reflex relation could be experimentally manipulated by isolating the stimulus (S) and response (R), and the correlation between these terms was presented as a mathematical function: $R=f(S)$. Skinner understood correlation as the necessary joint appearance of events described in the function, that is, the response should always occur in the presence of a stimulus and it should never occur in its absence. Skinner also proposed some specific measures of this correlation: latency, threshold, after-discharge, and the ratio R/S. He suggested that the reflex relation should be defined through this set of measures.

Skinner (1931/1999) noted that whenever there was a change in one of these measures, the others also exhibited some sort of change. Thus, the use of a generic term to describe this set of changes seemed convenient. The word chosen by Skinner was "strength," which identifies the state of correlation. For example, if a reflex had a low threshold, short latency, prolonged after-discharge and large R/S ratio, it would be considered "strong," if on the other hand, it had high threshold, long latency, short after-discharge and small R/S ratio, it would be considered "weak" (Skinner, 1931/1999, p. 501).

The unit of analysis employed in behavioral studies implied that "causality" of behavior resided in preceding events. There was not, up to that time, an emphasis on the consequences of behavior. The measurement used was the "strength" of the

response, which was defined as a descriptive term for a set of changes identified by various measurements. However, in practice, "strength" corresponded to only one measurement: response rate (e.g., Skinner, 1932).

Skinner (1935/1999) proposed a division among different types of conditioning and a "pseudotype." He identified two types of conditioning: (1) Type I, which he later called operant; and (2) Type II, which he later named respondent. The pseudotype referred to relations which involved discriminations and were based on both types. When these relations were based on Type I, Skinner observed that they also maintained characteristics of Type II and other characteristics that had not been identified in any type.

The division created by Skinner (1935/1999) was criticized by Konorski and Miller (1937), after which Skinner (1937/1999) replied to their criticisms. In the latter, he changed the terminology employed to name the types of conditioning. Instead of Types I and II, Skinner started using the terms Type R and Type S. The types were based on the correlation between the reinforcing stimulus and the response (Type R) or another stimulus (Type S). In the 1937 article, for the first time Skinner employed the terms "operant" (for Type R) and "respondent" (for Type S).

Subsequently, Skinner (1938/1991) essentially maintained the same ideas presented in 1937. He advocated describing the reflex as a correlation between S and R, the only important property of which was the coincidence of the occurrence of the terms (functional relation). Thus, Skinner saw the reflex as an analytical unit, and the word "reflex" started to encompass not only Type S (respondent) but also Type R (operant) conditioning.

Punishment: definition and terminology

a) First definition and terminology: 1935

Skinner (1935/1999) offered his first definition of punishment. In this article, he distinguished Type I and Type II conditioning. The definition of the conditioned reflex required contingency to the reinforcing stimulus: in Type I, reinforcement was correlated with the response, and in Type II, it was correlated with another stimulus. Skinner presented the definition of conditioned reflex, the terms involved and the alterations in strength he had identified, for both Types. The following excerpt describes the analysis made for Type I:

> A conditioned reflex is said to be conditioned in the sense of being dependent for its existence or state upon the occurrence of a certain kind of event, having to

do with the presentation of a reinforcing stimulus. A definition which includes more than this simple notion will probably not be applicable to all cases. At almost any significant level of analysis a distinction must be made between at least two major types of conditioned reflex. These may be represented, with examples, in the following way (where S = stimulus, R = response, (R – S) = reflex, = "is followed by", and [→] = "the strength of" the inclosed reflex):

TYPE I

S_0	———	R_0	⟶	S_1	———	R_1
(A) lever	–	pressing	–	food	–	salivation, eating
(B) "	–	"	–	shock	–	withdrawal, emotional change

Given such sequence, where $[S_1 – R_1]$ is $\neq 0$, conditioning occurs as a change in $[S_0 – R_0]$ – an increase in strength (positive conditioning) in (*a*) a decrease (negative conditioning) in (*b*). (Skinner, 1935/1999, p. 525)

Skinner (1935/1999) defined negative conditioning as a decrease in strength. This definition (e.g., lever pressing producing a stimulus and a subsequent decrease in its strength) is similar to a functional definition of punishment (e.g., Azrin & Holz, 1966). It therefore is suggested that Skinner was describing the operation (presenting a negative reinforcer) and effect (decrease in reflex strength) of what is called "punishment" under the terminology of "negative conditioning." Regarding the differences between Type I and II, Skinner stated that "The significant change in Type I may be either an increase or a decrease in strength…" (p. 528) and that "In Type I, stimuli may be divided into two classes, positively and negatively conditioning, according to whether they produce an increase or decrease when used as reinforcement" (Skinner, 1999/1935, p. 528).

Therefore, Skinner (1935/1999) asserted the possibility of negative conditioning as the opposite, regarding the direction of the change achieved in the reflex strength, from positive conditioning. The stimuli involved were classified as "positive" or "negative" according to the direction of behavioral change perceived.

Thus, Skinner's first definition of punishment (negative conditioning) was symmetrically opposite to the definition of positive conditioning, because the only difference between them would be in the direction of change in reflex strength produced by negative or positive reinforcing stimuli.

b) Second definition and terminology: 1938

After defining negative conditioning in 1935, Skinner (1938/1991) returned to this theme, mentioning negative conditioning for the first time in *Behavior of Organism*, as follows:

> The requirements for conditioning are some considerable strength of $S^1 .R^1$ and the connection indicated by →►. The effect is a change in $[s.R^0]$, which may be either an increase or, possibly, a decrease. In the present example of pressing lever the strength may increase if S^1 is, for example, food, and it may decrease if it is, for example, a shock. There are thus two kinds of reinforcing stimuli — positive and negative. The cessation of a positive reinforcement acts as a negative, the cessation of a negative as a positive (Skinner, 1938/1991, p. 66).

Skinner (1938/1991) stressed the conditions for the occurrence of Type R conditioning and stated that the conditioning could be an increase or possibly a decrease in strength. He also added a footnote asking the reader to read the section on negative conditioning in Chapter 4, where he questioned how the decrease in strength would be produced in Type R conditioning.

Skinner's classification of reinforcing stimuli was as in 1935, but he recognized another possible operation: the removal of a positive or negative reinforcer. Moreover, he mentioned its effects: the removal of a positive reinforcer would act as negative, that is, it would produce a decrease in strength, while the removal of a negative reinforcer would act as positive and increase strength.

Skinner (1938/1991) focused specifically on negative conditioning, distinguishing it from other procedures which decreased response strength, and questioned its status:

> One kind of reinforcing stimulus in Type R apparently produces a decrease in the strength of the operant. If pressing the lever is correlated with a strong shock, for example, it will eventually not be elicited at all. The result is comparable with that of adaptation or extinction, but there is little excuse for confusing these procedures. The distinction between extinction and a decline in strength with 'negative' reinforcement rests upon the presence or absence of the reinforcement and should be easily made.
>
> The effect of a reinforcing stimulus such as shock in decreasing strength may be brought about either by a direct reduction in the size of the reserve or

> by a modification of the relation between the reserve and strength. Only in the former case should we speak of negative conditioning. The process would be the opposite of positive conditioning and could be described as a reduction in the reserve not requiring the actual expenditure of responses, as in the case of extinction. It is not clear, however, that a reduction of this sort actually occurs, at least when the change begins after previous positive conditioning rather than at the original unconditioned strength (Skinner, 1938/1991, p. 108)

A few points about this quote deserve comment. The first is the use of the autoclitic "apparently" to refer to the decrease in strength of the operant through the presentation of a negative reinforcer. With this autoclitic, Skinner (1938/1991) provided one more indication that he was questioning the effect of negative conditioning on the strength of the operant. The second point concerns nomenclature. Skinner adopted the term "negative reinforcement" to describe what he called "negative conditioning" in 1935. Because he also used the latter expression in his 1938 work, it can be asserted that these expressions were, at least partially, interchangeable.

The third point refers to the possibility of differentiating negative reinforcement from extinction in procedural terms. Skinner stated that the difference between these operations resided in the presence or absence of reinforcement, that is, extinction is the breaking of a relation between the response and the reinforcing stimulus established beforehand; thus, there is no reinforcement in this procedure. Negative reinforcement/conditioning, in turn, was defined in *The Behavior of Organisms* by the presentation of a negative-reinforcer stimulus, so reinforcement is present in this latter case.

The fourth point concerns the examination of the status of negative reinforcement/conditioning as opposed to positive reinforcement. For that, Skinner (1938/1991) interjected a concept that permeated all of *The Behavior of Organisms* and which seems to be crucial to his explanatory system: the reflex reserve (hereafter, *reserve*). In the above excerpt, Skinner declared that the effects of negative reinforcement/conditioning could be explained by the effects that this procedure has on the reserve.

To sum up, Skinner (1938/1991) addressed negative reinforcement/conditioning differently than he did in 1935. In the latter, he appeared to assume the possibility of decreasing strength through negative conditioning and this would be the opposite of positive conditioning. In 1938, he directly investigated this possibility. What was responsible for this change? One reason was his development of the

reserve concept and its relation to behavioral operations. These two topics are the subject of the next section.

Possible reasons for the change in addressing negative reinforcement/conditioning

It was suggested in the preceding section that there was a change in the way that negative reinforcement/conditioning was addressed in the texts of 1935 and 1938. Skinner recognized this change in his autobiography:

> I had first used the term "negative reinforcement", incorrectly, to mean "punishment". I had assumed, along with almost everyone else, that punishment was simply the opposite of reward. You rewarded people to make them more likely and you punished them to make them less likely to behave in a given way. In my paper on two types of conditioning I said that reinforcing stimuli may be positive or negative "according as they produce an increase or a decrease in strength." But "reinforcing" means "strengthening" and in the *The Behavior of Organisms* I began to hedge. Consequences produced a change in the strength of an operant, "which may be either an increase or, *possibly* [italics added], a decrease." I said that the strength of pressing a lever may increase if the consequence "is, for example, food and it may decrease if it is, for example, shock," but a footnote referred the reader to a later section on "negative reinforcement" called merely "The Possibility of Negative Conditioning." Elsewhere I put the term "negative reinforcement" in quotation marks and questioned whether "a reduction of this sort actually occurs." My experiments had seemed to indicate that there was no effect on the reserve. (Skinner, 1979, p. 321)

In this excerpt, Skinner (1979) acknowledged his use of the term "negative reinforcement" to refer to punishment, as has already been noted in this review. Regarding the change in the use of negative reinforcement/conditioning in 1935 and 1938, these statements support the argument that was presented. They indicate that the argument appears to be correct regarding the reasons that led Skinner to change the meaning of negative reinforcement/conditioning. In this section, two possible reasons for this change will be discussed: (1) the development of the reserve concept and experiments published in 1938 and (2) criticism from Kornoski and Miller (1937).

a) The development of the "reserve of reflex" concept

The concept of the reserve has a history of formulation that spanned the 1930s and has a close relation to the concept of extinction, more specifically with the concepts of resistance to extinction and the extinction rate (Sério, 1990). While studying extinction, Skinner (1933a) noticed that the effects of conditioning went beyond the experimental hour. It was possible to observe an immediate change in the response rate while the conditioning occurred, but there was also a change that occurred after the conditioning (i.e., extinction). Skinner, then, encountered two possible measurements of the effects of conditioning: its immediate strength and its resistance to extinction. These measurements were systematically identified in a later article:

> We have distinguished elsewhere between the *immediately observed strength* of a conditioned reflex and *resistance to extinction*. The former is evaluated from some quantitative aspect of the reflex at a given elicitation, while the latter is inferred from the proprieties of the extinction curve subsequently obtained. There is no simple relation between them. Under repeated reinforcement, for example, a reflex will continue to develop resistance to extinction after its strength has reached an effectual maximum (Skinner, 1933b, p. 420)

There was not a simple relation between these two measurements, since it was possible that no change occurred in one measurement, for example, in the immediate strength, while conditioning was in force, but there were changes in the other measurement, for example, the resistance to extinction. Thus, Skinner (1933b) started noticing that conditioning produced at least two different effects on the organism and that it was necessary to deal with both.

Although Skinner (1933a, 1933b) discerned the need to address both effects of conditioning, he assumed that resistance to extinction was the appropriate measurement for his analysis because it showed the alterations produced by that environmental manipulation when the reinforcing stimulus was no longer present (extinction). The resistance to extinction clearly showed the effects of conditioning on behavior. Skinner (1933c) investigated the number of responses emitted in an extinction curve, given a conditioning number 1 (Nc=1), that is, one response reinforced. This manipulation led to the concept of the extinction ratio, which was the number of responses in extinction, given a reinforcement (Ne/Nc).

Skinner (1933a, 1933b, 1933c) was concerned with the number of responses that would be elicited after conditioning. According to Sério (1990), this directed Skinner to provide a new definition of conditioning[1] based on the number of responses that appeared in the extinction curve. With this new definition of conditioning and the concepts of resistance and extinction ratio, Skinner (1936) was equipped with a theoretical framework that permitted the presentation of the reserve concept:

> It has already been pointed out that the extinction curve is the proper measure of the effect of conditioning (8). Conditioning may be described as the creation of a certain number of potential responses which are later to be observed without further reinforcement. The number contributed to the total reserve by one reinforcement is the extinction ratio (7), which varies with the kind or condition of reinforcement (10,11). According to this view the elicitation of a response without reinforcement simply subtracts one from the number in reserve, although it remains to be shown, of course, that the effect of a failure to reinforce is constant throughout the curve. (Skinner, 1936, p. 308-309)

The definition of conditioning in this text was improved (Sério, 1990), because it began to be the number of *potential* responses that could be emitted during extinction, and the extinction ratio corresponded to the number of responses that were added to the total reserve by a reinforcement. Therefore, conditioning started to be addressed as an operation that created a reserve. In 1938, the concept of the reserve was finally formulated and appeared in nearly all chapters of the work, even in the discussion of the rejection of negative conditioning as a process opposed to positive conditioning.

Skinner (1938/1991) defined the reserve as the available activity that was created, concerning Type R, through conditioning. Skinner suggested, that the reserve was a hypothetical entity that had no physiological or local property in the organism. Hence, it was only a convenient way to aggregate certain experimental facts. MacCorquodale and Meehl (1948) observed that the concept of reserve in Skinner's work could be interpreted according to their definition of "intervening variables," that is, as constructs that are merely abstraction of empirical relations.

---

1 Skinner (1979) observed that if it were not for Pavlov, Magnus or Sherrington's influence, he would have considered response rate to have been his "basic data," however due to his knowledge of reflex theory, he wanted rate to be a measure of the strength of reflex.

Skinner (1938/1991) stated that the strength of the reflex was proportional to the reserve. Thus, there would be two ways to change it: (1) modifying the size of the reserve itself, or (2) modifying the proportionality between the reserve and strength. Skinner classified the behavioral operations in terms of their effects on the reserve: operations that involved elicitation would change the size of the reserve directly through conditioning (increase) and through extinction and fatigue (decrease). Other operations that produced effects on a set of reflexes would not modify the size of reserve, but rather the proportion between the reserve and strength through facilitation and some types of emotion (increase), inhibition and other types of emotion (decrease), and drive (increase or decrease). These operations would change the elicited rate of response, but not the number of responses available for elicitation. So, the concept of the reserve enabled the grouping of certain operations according to their effects on the reserve.

Analyzing the definition of negative reinforcement/conditioning proposed by Skinner (1938/1991), it can be seen how he would conclude that negative reinforcement/conditioning had effects opposite those of positive conditioning if it directly diminished the size of the reserve. Nevertheless, Skinner observed that it was not clear that this direct decrease occurred. In this manner, the only remaining alternative to explain the change in strength due to negative reinforcement/conditioning would be a change in proportion between strength and reserve. Thus, the explanatory mechanism of the decrease in strength generated by the negative reinforcement/conditioning should be confined to drive or emotion. Skinner opted to explain these effects as emotional, as will be discussed later.

Some essential characteristics of the reserve concept appeared in 1933, but it was only in 1936 that the concept came out in a clearer form. The change in the use of negative conditioning occurred between 1935 and 1938. The final (1938) elaboration of the reserve concept contributed directly to this change: from 1936 onwards, the explanation of behavioral operations began being based on the effects on the reserve and, with this rationale, it is contended that the experiments described below provided support to Skinner's position regarding negative reinforcement/conditioning; the following experiments were conducted with the concept of reserve as a guide.

Skinner (1938/1991) described five experiments that evaluated the effects of negative reinforcement/conditioning on the reserve: (1) in extinction (Experiments I and II); (2) in alternation with positive reinforcement (Experiments IIIA and III B); (3) in extinction after a history of exposure to negative and positive

reinforcement (Experiment IV); and (4) in extinction after adaptation to negative reinforcement (Experiment IV). In all experiments, periodic reinforcement was employed, most likely a fixed-interval 4-min schedule[2], and a "slap" generated by the reverse movement of the bar when pressed was used as aversive stimulus.

The way the experiments were outlined and how the data were explained (see Skinner, 1938/1991 for further details) has the reserve concept as a basis. The effects of negative reinforcement/conditioning were tested in most experiments (I, II, IV and V) in extinction, which reflects Skinner's (1938/1991) commitment to extinction (or resistance to extinction) as a measure of conditioning. If any of the procedures had had direct effects on the reserve, they would have been revealed during extinction (curve and total of responses emitted). Experiments on negative reinforcement/conditioning demonstrated that there was no effect on the *size* of the reserve; in other words, there was no change in the total number of responses available for emission. The changes in response strength were, according to him, explained by the change in the factor proportionality between reserve and strength.

b) The criticism of Konorski and Miller (1937)

Skinner (1935/1999) was criticized by Konorski and Miller (1937) for his construction of conditioning Type I and, when mentioning the description of negative conditioning in Skinner (1935/1999), for the supposition that negative conditioning merely diminished the strength of the reflex. The authors presented a formulation based on (1) conditioning of response properties into noxious stimuli, and (2) the emergence of an incompatible reflex. Skinner (1937), in a reply to their criticism dedicated a few lines to negative conditioning, but seems not to agree with the formulation put forward by them:

> It is essential in its kind of formulation that one reflex be considered at a time since our data have dimensions of changes in reflex strength. The development of an antagonistic response when a reinforcement in Type R is negative requires a separate paradigm, either Type R or Type S (Skinner, 1937/1999, p. 542).

---

2 Skinner (1938/1991) always mentioned periodic conditioning, however he did not specify, except in Experiment IIIA, the interval duration. It is suggested that the schedule used by Skinner in the experiments was FI 4 min, because in Estes's punishment monograph (1944), on which Skinner was the research advisor, the schedule is FI 4 min. Holland and Skinner (1961), in presenting Experiment II, published in 1938, also described the schedule as FI 4 min.

Skinner (1937/1999) disagreed on the need to introduce another paradigm to explain negative reinforcement/conditioning by invoking the emergence of an antagonist response. It would be one more reflex to be handled, which, to Skinner, would disturb the analysis in terms of modifications in strength. In dealing with one more simultaneous reflex, the alterations in strength would not be directly measurable because they would be the product of the appearance of this new reflex. As a result, they would only be discernable indirectly.

Although Skinner (1937/1999) had not fully adopted the formulation proposed by Konorski and Miller (1937), at least one element of it could be identified in 1938, when he, while demonstrating that negative reinforcement/conditioning did not change the size of reserve, opted to explain its effects in terms of alteration in the proportionality between strength and reserve through emotion.

Explanatory mechanisms of behavioral suppression

As previously noted, of the two ways to change response strength — the size of the reserve or the proportionality between reserve and strength — Skinner (1938/1991) concluded that the most appropriate explanatory mechanism for negative reinforcement/conditioning was the latter. Given the operations that changed the proportionality between reserve and strength, Skinner opted, as has been suggested, for emotion:

> The alternative case of a modification between the strength and the reserve comes under the heading of emotion as defined later. The emotional reaction to the shock is conditioned according Type S in such a way that the lever or incipient movements of pressing the lever become a conditioned stimulus capable of eliciting it. The effect of the emotional state is to reduce the strength of the response. Responses are not made when the lever is presented, not because there are no responses in the reserve, but because the lever sets up an emotional state in which the strength is depressed. The resulting failure to respond is obviously related to the phenomenon of repression. (Skinner, 1938/1991, pp. 108-109)

In this assertion, Skinner (1938/1991) described how a stimulus (e.g., shock) could alter the strength of an operant through an emotional state, not by changing the number of responses emitted in extinction, but, rather, by changing the behavioral flow. Thus, Skinner adopted one of the elements from the criticism by Konorski and Miller (1937) to explain negative reinforcement/conditioning: the aversive

conditioning of properties of the response itself (e.g., incipient movements of bar pressing) through pairing with shock.

Nevertheless, Skinner (1938/1991) extended the aversive conditioning to properties of the experimental situation—the bar—which functioned as a discriminative stimulus for the pressing response. Thus, it was observed that a conflict is produced by one single environmental event being correlated with stimuli of distinct functions. Another element from Konorski and Miller's formulation (1937), the production of an antagonist reflex, was not used by Skinner (1938). The response suppression generated by negative reinforcement/conditioning would be, in Skinner's view (1938/1991), a product of an emotional state generated by conditioned stimuli: incipient movements and the bar.

The emotional state generated by presenting a negative reinforcing stimulus was used by Skinner (1938/1991) to explain the response suppression observed during and after negative reinforcement/conditioning. The use of the concept of an emotional state could lead to the interpretation that Skinner was appealing to a hypothetical construct, something that has no experimental dimension, and which would be used as a "cause" of the strength alterations. However, the term "emotional state," just as "drive," was employed as an intervening variable, as described by MacCorquodale and Meehl (1948), to operations that changed groups of reflexes and/or was not unique in its effects (Skinner, 1938/1991).

For example, distinct operations could have the same effect on a group of reflexes: presenting a shock or a loud noise and failing to present the food may have the same effect on the strength of the reflexes of ingestion (decrease) and bar pressing (decrease). Thus, Skinner (1938/1991) appealed to the "emotional state" as a term which gathered similar modifications in a set of reflexes. Then, Skinner's explanation of negative reinforcement/conditioning is represented in Figure 1.

The bar pressing response generates shock. The shock is paired with the bar and also with response proprieties directed toward the bar. The operation "presentation of shock" produces an emotional state that correspond to the alterations in strength in more than one reflex (the reflex of bar pressing and that of ingestion, for example). After repeated exposures to the experimental arrangement, the bar and movements toward it (proprieties of the response) start to produce, due to the pairing with shock, the same emotional state, identified as a decrease in the strength of more than one reflex.

A question is that if emotion, as well as drive, induced changes in groups of reflexes and modified the proportionality between reserve and strength, why did

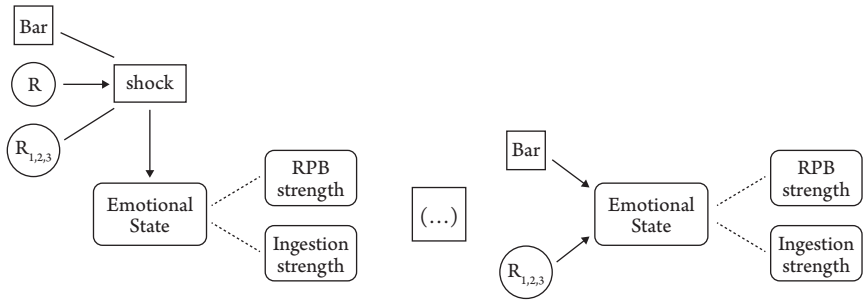

Figure 1. Schematic representation of the explanatory mechanism for negative reinforcement/conditioning in Skinner (1938/1991). Notations: R = response; $R_{1,2,3}$ = proprieties of response; arrow = produce; continuous line = paring; dotted line = stands for; (...) = exposure to the arrangement, RPB = bar pressing response.

Skinner (1938/1991) explain the effects of negative reinforcement/conditioning through emotion and not through drive. He recognized that the difference between drive and emotion is subtle and, many times, difficult to establish. Both were identified with alterations in reflex strength; however, one possible difference was suggested.

This difference concerns compensatory effects or the response recovery: "As will be pointed out in Chapter Eleven, Emotion and Drive are closely related phenomena, but it will be shown in Chapter Ten that a reduced rate due to lowered drive is not compensated for subsequently" (Skinner, 1938/1991, p. 157). This distinction is more plausible for justifying Skinner's emotion-based explanation of negative reinforcement/conditioning. While describing the experiments on negative reinforcement/conditioning, Skinner (1938/1991) highlighted the compensatory effect, especially in Experiments II and III, and an adaptation effect (Experiments III and IV). These data probably served as the basis for Skinner identifying the response suppression produced as an emotional effect and not as a drive effect because compensatory increases were not characteristic of drive changes.

Symmetry and asymmetry in relation to positive reinforcement in the 1930s

The secondary literature on punishment brings up two theoretical views on this phenomenon: symmetry and asymmetry in relation to the reinforcement. Authors frequently referred to in regard to this debate are Azrin and Holz (1966) and Skinner (1953/2005), respectively (Holth, 2005). As previously noted, it seems to be

believed that generally the symmetrical position prevails among behavior analysts (Michael,1975; Skiba & Deno, 1991; Lerman & Vondran, 2002; Gongora, Mayer & Mota, 2009), but asymmetry remains as an alternative position (Hineline & Rozales-Ruiz, 2013).

Understanding punishment as asymmetrical to reinforcement means considering that its effects are not exactly opposite to positive reinforcement and that the behavioral mechanisms of each are different (Spradlin, 2002). Subsequently, punishment would not directly decrease the frequency of a response. Even though a decrease in frequency may occur, such a decrease is the product of other behavioral processes. Thus, punishment would not have the status of a primary behavioral process.

In the 1930s, Skinner's treatment of punishment (negative reinforcement/conditioning) was changed, in 1935 and again in 1938. In 1935, the definition of negative reinforcement/conditioning reflected the symmetrical position, because Skinner (1935/1999) presented negative conditioning as a type of conditioning the effect of which was to decrease reflex strength, in contrast to positive conditioning. It is noteworthy that Skinner did not appeal to any other explanatory mechanism to encompass the effects of negative conditioning.

In 1938, Skinner took a completely different theoretical posture. Positive conditioning was classified as one of the operations that changed the size of reserve, i.e., increased it. The operation symmetrically opposite to it was extinction, which modified the size of the reserve, i.e., decreased it. "The important thing is the process of conditioning and its reciprocal process of extinction" (Skinner, 1938/1991, p. 61). One thing that draws attention in this statement is the use of the word "reciprocal"; that is, extinction was considered as opposite process to positive conditioning.

If positive conditioning was the creation of a potential number of responses that would be available for emission, then extinction was responsible for exhausting the responses (Skinner, 1938/1991). When classifying the operations in terms of their effects on the reserve, Skinner did not talk about negative reinforcement/conditioning. This position makes sense when one understands that the explanatory mechanism of this phenomenon was emotion. In view of this, negative reinforcement/conditioning did not have a place among the dynamic laws.

It has been verified in the present review that Skinner (1938/1991) (1) did not consider negative reinforcement/conditioning as an opposite process to positive conditioning and (2) explained its effects through another behavioral mechanism (emotion). These characteristics clearly identify the Skinnerian position in 1938

as asymmetrical. The reserve concept, being a hydraulic model and based on an input-output notion, probably influenced this type of position, because symmetrical would require opposite but equal effects on the reserve, i.e., the addition and subtraction of responses. In this case, the operations that met these criteria were positive conditioning and extinction.

Skinner (1938/1991) also contributed to the discussion related to the intensity of stimuli in negative reinforcement/conditioning. In Experiment I (p.151-155), using a prolonged bar slap, the curve obtained seemed to indicate that negative reinforcement/conditioning subtracted, similar to extinction, responses from reserve. However, Skinner did not find the same result when he used a brief slap. He could have explained the difference according to the type of stimulus used. He could have said that a strong stimulus subtracted responses from the reserve and that a weak stimulus would not work in the same way. However, he used emotion as an explanation for both situations, which is understandable because he selected the explanation that could be used for both situations (conceptual economy).

Another point to be highlighted is the notion of suppression. Skinner (1938/1991) reserved the term suppression for operations that altered the proportionality between the reserve and response rate: "The notion of suppression applies to any factor altering the relation between the reserve and the rate of responding in such way that the latter is reduced" (Skinner, 1938/1991, p. 102). Therefore, the notion of suppression was applied to emotion and drive and to the cases in which another reflex became prepotent. Thus, the use of this word explicitly supports an asymmetrical position, because otherwise, he would use the word "weakening," which was the word he used for the effect of extinction.

Final considerations

The terminology used by Skinner to refer to punishment was first negative conditioning (1935) and then negative reinforcement or negative conditioning (1938). The first definition of negative conditioning (1935) considered the process as opposite to positive conditioning, that is, decreasing reflex strength. The second definition (1938) stopped considering negative reinforcement/conditioning as the opposite of positive conditioning. The reasons for this change were the formulation of the concept of reserve, the experiments published in 1938, and the critique of Konorski and Miller (1937).

Here, it is suggested that in 1935 the definition of negative conditioning would be considered symmetrical relative to positive reinforcement, and that in 1938 the

definition would be considered asymmetrical. The reserve concept seems to have been crucial to this kind of division, because the behavioral operations were separated according to it.

The end of the 1930s brought a change in Skinner's system. The reserve concept, which seems to have been one the core concepts of Skinner's explanation of behavior (1938/1991), was jeopardized. Skinner (1940) presented data, different from the data presented in 1938, relating the reserve to drive. He observed that variations in drive produced different extinction curves across a set of reinforcement conditions; that is, the number of responses was not constant, which suggested to Skinner problems with the reserve concept. This 1940 paper marks the beginning of his abandonment of the reserve concept, although he was not yet willing to do so explicitly, as he affirmed later (Skinner, 1979).

In a letter sent to Michael Zeiler (unpublished letter from Skinner to Michael Zeiler, 1977), Skinner affirmed that the development of new schedules of reinforcement made the notion that a given number of responses would appear without reinforcement meaningless. The study of increasingly complex scheduling arrangements made the concept of the reserve unnecessary because they broke the input-output relation predicted by the concept of the reserve.

Skinner's behavioral system in the 1930s was, from the middle of the decade, completely based on the reserve concept, including the rejection of negative reinforcement/conditioning as the opposite of positive conditioning. Because this concept was abandoned, the question arises as to how Skinner addressed negative reinforcement/conditioning in subsequent decades. Secondary sources that have analyzed Skinner's (1953) treatment of punishment affirmed that he treats negative reinforcement/conditioning (in this moment officially called "punishment") as asymmetrical to positive reinforcement (e.g., Hotlh, 2005; Mayer, Gongora & Mota, 2009; Carvalho Neto & Mayer, 2011). This position raises questions because according to the analysis of his position during the 1930s, the core for the asymmetrical thesis (reserve) was abandoned.

Therefore, we can ask how Skinner kept the asymmetrical position after the 1930s without the notion of the reserve. Which conceptual tools aided this preservation? This review reveals that the historical understanding of the concept of punishment in the 1930s, linking it to other concepts inside Skinner's body of work, generates a path to issues that must be pursued when studying the concept in subsequent years. In the 1930s, the terminology was different, there were changes in the definition and explanation and the concept of reserve, the basis for the asymmetry

position, was abandoned. In reading the formulation of punishment in the following decades, primarily in 1953, these points should guide the reader to inquire: Why was this terminology chosen? Have the definitions and explanations remained the same? How can you think about punishment asymmetrically without the concept of reserve? The answers to these questions are the topic of the companion article to this in the following issue.

References

Araujo, S. F. (2016). A investigação histórica de teorias e conceitos psicológicos: breves considerações metodológicas [The historical investigation of theories and psychological concepts: brief methodological considerations]. In C. Laurenti, C. E. Lopes & S. F. Araujo (Orgs.), *Pesquisa Teórica em Psicologia: Aspectos Filosóficos e Metodológicos*. São Paulo: Hogref.

Azrin, N. N., &Holz, W. C. (1966). Punishment. In W. K. Honig (Org.), *Operant behavior: Areas of research and application* (pp. 380-447). Englewood Cliffs: Prentice-Hall.

Carvalho Neto, M. B., & Mayer, P. C. M. (2011). Skinner e a assimetria entre reforçamento e punição [Skinner and the asymmetry between reinforcement and punishment]. *Acta Comportamentalia, 19*, 21-32.

Catania, C. A. (1999). *Aprendizagem: comportamento, linguagem e cognição [Learning: behavior, language and cognition].* (4ª ed.; D. G. Souza et al., Trads). Porto Alegre, RS: Artmed.

Dinsmoor, J. A. (1954). Punishment: I. The avoidance hypothesis. *Psychological Review, 61*, 34-46.

Dinsmoor, J. A. (1955). Punishment: II: An interpretation of empirical findings. *Psychological Review, 62*, 96-105.

Dinsmoor, J. A. (1977). Escape, avoidance and punishment: where do we stand? *Journal of the Experimental Analysis of Behavior, 28*, 83-95.

Dinsmoor, J. A. (1998). Punishment. In W. T. O'Donahoe (Ed.), *Learning and Behavior Therapy* (pp. 188-204). Needham Heights, MA: Allyn & Bacon.

Estes, W. K. (1968). An experimental study of punishment. In E. E. Boe & R. M. Church (Eds.), *Punishment: Issues and Experiments* (pp. 108-165). New York, NY: Appleton-Century-Crofts. (Original work published in 1944).

Gongora, M. A. N., Mayer, P. C. M., & Mota, C. M. S. (2009). Construção terminológica e conceitual do controle aversivo: período Thorndike-Skinner e algu-

mas divergências remanescentes [Terminological and conceptual construction of aversive control: Thornidike-Skinner period and remaining divergences]. *Temas em Psicologia, 17*, 209 – 224. Retrived from: http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1413389X2009000100017&lng=pt&tlng=pt.

Griffin, J. C., Paisey, T. J., Stark, M. T., & Emerson, J. H. (1988). B. F. Skinner's position on aversive treatment. *American Journal of Mental Retardation, 93*, 104-105.

Hineline, P. N. (1984). Aversive Control: a separate domain? *Journal of Experimental Analysis of Behavior, 42*(3), 495-509. doi: 10.1901/jeab.1984.42-495.

Hineline, P. N., & Rozales-Ruiz, J. (2013). Behavior in relation to aversive events: punishment and negative reinforcement. In G. J. Madden, W. V. Dube, T. D. Hackenberg, G. P. Hanley, & K. A. Lattal (Eds.), *APA Handbook of Behavior Analysis* (pp. 483-512). Washington, DC: American Psychological Association.

Holland, J. G., & Skinner, B. F. (1961). *The Analysis of Behavior*. New York, NY: McGraw-Hill.

Holth, P. (2005). Two definitions of punishment. *The Behavior Analyst Today, 6*(1), 43-47.doi:http://dx.doi.org/10.1037/h0100049.

Konorski, J., & Miller, S. (1937). On two types of conditioned reflex. *Journal of General Psychology, 16*, 264-272. Retrived from: http://psychclassics.yorku.ca/Skinner/Konorski/

Lerman, D. C., & Vorndran, C. M. (2002). On the status of knowledge for using punishment: implications for treating behavior disorders. *Journal of Applied Behavior Analysis, 35*(4), 431-434. doi: 10.1901/jaba.2002.35-431.

MacCorquodale, K., &Meehl, P. E. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychological Review, 55*(2), 95-107. doi: dx.doi.org/10.1037/h0056029.

Martins, T. E. M., Carvalho Neto, M. B., & Mayer, P. C. M. (2013). B. F. Skinner e o uso do controle aversivo: um estudo conceitual [B. F. Skinner and the use of aversive control: a conceptual study]. *Revista Brasileira de Terapia Comportamental e Cognitiva, 15,* 5-17.

Mayer, P. C. M., &Gongora, M. A. N. (2011). Duas formulações comportamentais de punição: definição, explicação e algumas implicações [Two behavioral formulations of punishment: definition, explanation and some implications]. *Acta Comportamentalia, 19*, 47-63.

Michael, J. (1975). Positive and negative reinforcement, a distinction that is no longer necessary; or a better way to talk about bad things. *Behaviorism, 3,* 33-44.

Morris, E. K., Todd, J.T., Midgley, D., Schneider, S. M., & Johnson, L. M. (1990). The History of Behavior Analysis: Some historiography and a bibliography. *The Behavior Analyst, 13*(2), 131-158. Retrieved from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2733434/.

Rampolla, M. L. (2015). *A pocket guide to writing in history* (8thed.). Boston/New York: Bedford/St Martin's.

Sério, T. M. A. P. (1990). *Um caso na história do método científico: do reflexo ao operante [A case of history in scientific method: from reflex to operant]* (Unpublished doctoral dissertation). Pontifícia Universidade Católica de São Paulo, São Paulo, SP.

Sério, T. M. A. P. (1993). Relendo B. F. Skinner e aprendendo com ele [Reading B. F. Skinner and learning with him]. *Acta Comportamentalia, 2,* 155-166.

Sidman, M. (1995). *Coerção e suas implicações [Coercion and its fallout]* (M. A. Andery& T. M. Sério, Trads). Campinas: Editorial Psy. (Original work published in 1989).

Skiba, R. J., & Deno, S. (1991). Terminology and behavior reduction: the case against punishment. *Exceptional Children, 57*, 298-316.

Skinner, B. F. (1932). Drive and reflex strength. *Journal of General Psychology, 6*, p. 22-37.

Skinner, B. F. (1933a). On the rate of extinction of conditioned reflex. *Journal of General Psychology, 8*, 114-129.

Skinner, B. F. (1933b). "Resistance to extinction" in the processes of conditioning. *Journal of General Psychology, 9*, 420-429.

Skinner, B. F. (1933c). The rate of establishment of a discrimination. *Journal of General Psychology, 9*, 302-50.

Skinner, B. F. (1936). Conditioning and extinction and their relation to drive. *Journal of General Psychology, 14*, 296-317.

Skinner, B. F. (1940). The nature of operant reserve. *Psychological Bulletin, 37*, 423.

Skinner, B. F. (1977, November 10). Unpublished letter to Michael Zeiler. Harvard University Archives: Cambridge, MA.

Skinner, B. F. (1979). *The shaping of a behaviorist.* New York, Knopf.

Skinner, B. F. (1938/1991). *The behavior of organisms: An experimental analysis* (Rev. ed.). Acton, MA: Copley Publishing Group. (Original work published in 1938).

Skinner, B. F. (1999). The concept of reflex in description of Behavior. In V. G. Laties & A. C. Catania (Eds.), *Cumulative Record: Definitive Edition* (pp. 475-503). Acton, MA: Copley Publishing Group. (Original work published in 1931).

Skinner, B. F. (1999). Two types of conditioned reflex and a pseudo-type. In V. G. Laties & A. C. Catania (Eds.), *Cumulative Record: Definitive Edition* (pp. 525-534). Acton, MA: Copley Publishing Group. (Original work published in 1935).

Skinner, B. F. (1999). Two types of conditioned reflex: A reply to Konorski and Miller. In V. G. Laties & A. C. Catania (Eds.), *Cumulative Record: Definitive Edition* (pp. 535-5543). Acton, MA: Copley Publishing Group. (Original work published in 1937)

Skinner, B. F. (2005). *Science and human behavior*. New York, NY: The B. F. Skinner Foundation (Original work published in 1953).

Solomon, R. L. (1964). Punishment. *American Psychologist, 19*, 239-253.

Spradlin, J. E. (2002). Punishment: a primary process? *Journal of Applied Behavior Analysis, 35*(4), 475-477. doi: 10.1901/jaba.2002.35-475.

Thorndike, E. L. (1931). *Human Learning*. Cambridge: The M.I.T. Press.

Viera Pinto, A. (1979). *A ciência e a existência: problemas filosóficos da pesquisa científica* [*Science and existence: phylosophical problems of scientific research*] (2ª ed.). Rio de Janeiro: Paz e Terra.

Recibido Noviembre 22, 2018 /
Received November 22, 2018
Aceptado Agosto 27, 2019 /
Accepted August 27, 2019