Conditioned Reinforcement and Backward Association

Arthur Prével[1], Vinca Rivière[1], Jean-Claude Darcheville[1], and Gonzalo P. Urcelay[2]

[1]Laboratoire SCALab

UMR 9193

Université de Lille Nord de France

Campus de Lille 3

Domaine Universitaire du Pont de Bois

B.P. 60149

59653 Villeneuve d'Ascq Cedex

France

[2]Department of Neuroscience, Psychology and Behaviour

University of Leicester

Centre for Medicine, Lancaster Road,

Leicester,

LE1 7HA

UK

Correspond with:

Arthur Prével

Email: arthur.prevel@etu.univ-lille3.fr

Manuscript in press: *Learning & Motivation.*

Abstract

In the present study, excitatory backward conditioning was assessed in a conditioned reinforcement paradigm. The experiment was conducted with human subjects and consisted of five conditions. In all conditions, US reinforcing value (i.e. time reduction of a timer) was assessed in phase 1 using a concurrent FR schedule, with one response key leading to US presentation and the other key leading to no-US. In phase 2, two discrete stimuli, S+ and S-, were paired with US and no-US respectively using an operant contingency. For three groups, backward contingencies were arranged, and two of these were designed to rule out a trace (forward) conditioning interpretation of the results. The two other groups served as control conditions (forward and neutral conditions). Finally, in phase 3 for all groups the CSs were delivered in a concurrent FR schedule similar to phase 1, but with no US. Responding during phase 3 showed conditioned reinforcement effects and hence excitatory backward conditioning. Implications of the results for conditioned reinforcement models are discussed.

Keywords: Backward conditioning; Conditioned Reinforcement Hypothesis; Delay Reduction Theory; Humans; Signal Hypothesis

1. Introduction

Conditioned reinforcement is an old concept in the study of animal and human behavior. Its core idea is that an initially neutral stimulus (NS), because of its pairing with a primary reinforcer, acquires the capacity to serve as an effective reinforcer. Here, by reinforcement we mean the increase in the frequency of an operant behavior by the contingent presentation of a stimulus. Early evidence of conditioned reinforcement was observed in studies conducted in chimpanzees by Wolfe (1936) and Cowles (1937), with the use of tokens as conditioned reinforcers, or in a study conducted in rats by Bugelski (1938) studying resistance to extinction. But perhaps the most representative demonstration of conditioned reinforcement was the study conducted by Skinner in 1938, with the new response procedure. In Skinner's experiment, the sound of a pellet dispenser was first paired with the delivery of food without the requirement of a response by the rats (i.e. stimulus-stimulus pairing). During a second phase, a lever was introduced in the chamber and pressing the lever (i.e. the new response) produced the sound of the pellet dispenser (without food delivery). Evidence of conditioned reinforcement was shown by an increase in lever press frequency with the contingent delivery of the sound.

Multiple procedures have been designed since the discovery of conditioned reinforcement, and these procedures can be broadly divided in two categories (Williams, 1994). In the first category, the conditioned reinforcer is isolated from the primary reinforcer after the initial pairings and presented contingent on some behavior. The most significant example is the new-response procedure proposed by Skinner (1938) cited above (see also Sosa, dos Santos, & Flores, 2011). Another example is the resistance to extinction procedure (Bugelski, 1938; Urushihara, 2004), where resistance to extinction of an operant response is increased by the contingent presentation of a conditioned reinforcer. This effect is, for example, demonstrated with a comparison group where the operant response is not followed by the primary or the conditioned reinforcer, and where the resistance to extinction is reduced.

In the second category designed for the study of conditioned reinforcement, the conditioned stimulus (CS) is also made contingent on an operant response but the pairings with the primary reinforcer are maintained to avoid Pavlovian extinction. Well known procedures developed in this category are the chain and concurrent-chains schedules (Fantino, 1977; Kelleher & Gollub, 1962). In a concurrent-chains schedule, two concurrent initial-link schedules (e.g. VI 120 conc. VI 120) produce the transition to mutually exclusive terminal-link schedules (e.g. VI 30 and VI 90) producing the primary reinforcer. Transition from initial-link to terminal-link is signaled by different stimuli (e.g. red light and green light) and these stimuli are assumed to develop conditioned reinforcing properties. A second well-known procedure is the observing response procedure (Dinsmoor, 1983; Shahan and Podlesnik, 2005, 2008). In this procedure, an un-signaled reinforcement schedule delivering food alternates with extinction (i.e. a mixed schedule) on one response key. On a second response key (i.e. the observing key), a brief stimulus presentation is associated with the reinforcement schedule (S+) and a second stimulus is associated with the extinction schedule (S-). These stimuli are produced by pressing on the observing key, and responding on the observing key is supposed to be maintained by the conditioned reinforcement properties of S+. Finally, a last procedure known in this second category is the token procedure (Hackenberg, 2009), where tokens are earned and exchanged for accesses to primary reinforcers. Here tokens are supposed to act as conditioned reinforcers.

Most of the experiments on conditioned reinforcement using the procedures described above have been conducted in rats and pigeons. However, there have also been reports of conditioned reinforcement in human subjects. For example, evidence of conditioned reinforcement in a free operant situation was found with psychiatric patients (Levin & Sterner, 1966) and with children (Myers & Myers, 1962), and numerous papers have reported the effect of human attention in the increase of appropriate behaviors (Hall, Lund & Jackson, 1968; Jones, Drew, & Weber, 2000; Northup, Broussard, Jones, & Herring, 1995). Furthermore, chain schedules have been implemented in children (Long, 1963), and concurrent chain schedules have been studied in adults using a video-game (Leung, 1989, 1993). More recently, the observing response procedure (Fantino & Silberberg, 2010) was studied using a video-game. Finally, studies have also reported conditioned reinforcement effects by tokens in a token economies procedure (Kazdin, 1977).

As discussed above, early evidence of conditioned reinforcement was found in the first half of the 20th century, and a large literature has evolved. But the mechanisms underlying conditioned reinforcement are still debated in the literature (Shahan, 2010), and there is no consensus on what drives conditioned reinforcement. In the following sections, we will review some of the most important hypotheses on conditioned reinforcement.

One of the most influential hypotheses is the conditioned reinforcement hypothesis, or CRH (Dinsmoor, 1983; Skinner, 1938, 1953). CRH was developed in a time dominated by S-R learning theories (Hull, 1943; Spence, 1950) and hence was influenced by this framework. Its core idea is that a neutral stimulus (NS) will develop its own reinforcing value because of its pairing with a primary appetitive stimulus. In other words, a stimulus will develop the capacity of strengthening a stimulus-response association because of previous stimulus-stimulus pairings with the primary reinforcer. Conditioned reinforcement was an influential concept for theorists working in the S-R framework because it permitted them to translate the results from labs to natural situations. Different forms of the CRH were developed (see for example Kelleher & Gollub, 1962, for a review), but the core idea of CRH is that the temporal contiguity between the conditioned stimulus (CS) and the appetitive unconditioned stimulus (US) is an important variable in the development of appetitive value by the CS, even a necessary and sufficient condition (Skinner, 1953). The influence of CS-US delay was demonstrated for example by Bersh (1951) and Jenkins (1950), who showed that the number of lever presses made before CS delivery was reduced with an increased delay between the CS and US. CRH has more recently been supported by Donahoe and Palmer (2004) and Donahoe (2014).

A second well-known hypothesis on conditioned reinforcement is the Delay Reduction Theory, or DRT (Fantino, 2008; Fantino, Preston, & Dunn, 1993; Preston & Fantino, 1991). DRT was originally developed to explain choices in concurrent-chain schedules of reinforcement and the influence of the stimuli signaling the transition from initial to terminal links on response allocation. Although multiple forms of DRT have been developed, its core idea is that the effectiveness of a stimulus as a conditioned reinforcer may be predicted by its reduction in the length of time to primary reinforcement, measured from the onset of the conditioned reinforcer. In its simplest form, DRT may be stated by:

Reinforcement strength of stimulus A

$$f\left(\frac{T-t_A}{T}\right) \tag{1}$$

Where T is the averaged time between primary reinforcer presentations and t_A is the time between the conditioned reinforcer and primary reinforcer onset. So, DRT assumes that the more a conditioned reinforcer is correlated with reduction in waiting time to reinforcement, the more it will develop reinforcing properties. This effect was for example demonstrated in an experiment by Fantino (1969), where he found a large preference in a concurrent chain schedule paradigm for a VI 90 VI 30 schedule over a VI 30 VI 90 schedule. Finally, we may note the similarity between DRT and some models of conditioning, like Rate-Expectancy Theory (RET) model (Gallistel & Gibbon, 2000) or the recent Information-based model of conditioning (Gallistel, Craig, & Shahan, 2014).

Another important hypothesis in the study of conditioned reinforcement is the signal hypothesis, or SH (Shahan, 2010). SH was developed in a theoretical context influenced by the informational hypothesis (Egger & Miller, 1963; Kamin, 1969) where a purposive view of behavior (Tolman & Brunswik, 1935) was reconsidered. Different forms of SH were developed (Bolles, 1975; Davison & Baum, 2006; Longstreth, 1971; Schuster, 1969; see Shahan, 2010, for a review), but all share the idea that a stimulus which is predictive of a primary reinforcer will function as a signal (or signpost, means-to-an-end, etc.) for that primary reinforcer. Instead of a strengthening process, the appetitive CS is supposed to inform the subject about primary reinforcer presentation and influence the subject's behavior by this informational process. Also, it should be noted that the SH is similar to classical theories of associative learning (Mackintosh, 1975; Pearce & Hall, 1980; Rescorla & Wagner, 1972) because it assumes that the appetitive CS in a Pavlovian paradigm acts as a signal for the US.

The present experiment was designed to study the hypotheses reviewed above. In order to do that, we designed an experiment in a conditioned reinforcement paradigm using backward instead of a forward paired stimuli. To review, in backward conditioning, a CS is delivered *after* a US (rather than preceding it). Backward conditioning is usually interpreted as a pairing procedure that does not lead to excitatory conditioning, because of the lack of predictive relationship between the CS and US. But several studies conducted in Pavlovian paradigms (Ayres, Haddad, & Albert, 1987; Burkhardt, 1980; Chang, Blaisdell, & Miller, 2003; Heth, 1976; Mahoney & Ayres, 1976; Spetch, Wilkie, & Pinel, 1981; Spetch, Terlecki, Pinel, Wilkie, & Treit, 1982; Tait & Saladin, 1986) have shown that a backward CS can develop excitatory properties and control response elicitation, like freezing responses. These results have been interpreted as incompatible with the classical associative learning theories cited above, because these models suppose that the CS has to be a non-redundant predictor of the US to become an excitatory CS (Arcediano, Escobar, & Miller, 2003). Because the signal hypothesis of conditioned reinforcement is in some way similar to these models, it appears to us that a test of this hypothesis by assessing backward conditioned reinforcement is needed. Indeed, if the conditioned reinforcer has to be a predictor of the primary reinforcer to develop its reinforcing properties (i.e. increasing response rate), with a backward pairing we should not observe conditioned reinforcing effects by the CS. Moreover, in order to test DRT, we also included two supplementary backward conditions (see rationale below).

To our knowledge, only one study conducted by Urushihara (2004) has documented backward conditioned reinforcement. The study was conducted in rats using a resistance to extinction paradigm. In human subjects, studies conducted by Arcediano et al. (2003) have provided evidence of excitatory backward conditioning, but using a Pavlovian second-order conditioning paradigm. In addition to testing the hypotheses cited above, the current study is the first (to our knowledge) to assess conditioned reinforcement after backward pairings in human subjects. This experiment follows the criteria established by Spetch et al. (1981) for the demonstration of excitatory backward conditioning, notably the use of two control backward pairing procedures (B2 and B3 conditions; see below) to rule out conditioning between the CS presented on trial n and the US presented on trial n + 1. The implications of our results for theoretical models of conditioned reinforcement and associative learning are debated in the discussion section.

2. Materials and Methods

2.1 Subjects

Fifty human subjects (38 females and 12 males) were invited to participate in the experiment. They were 18 to 44 years old ($\acute{x}$ = 20.9 and σ = 4.7). Most of them were students at the Lille 3 University, and their informed consent was obtained before the experiment began. Ten participants were randomly assigned to each condition (5 conditions).

2.2 Apparatus

The experiment was conducted in an experimental room at the SCA-Lab laboratory and run on a PC using Windows 2000 operating system. The experiment was programmed on Matlab using the Psychophysics Toolbox extension. Subjects were seated on a chair at around 60 cm from the monitor. The US was reduction in the time taken to complete the experiment (Darcheville, Prével, & Rivière; unpublished results): the experiment began with the computer screen displaying a timer on which was written "180.00" (time in minutes). The timer was placed in the top right-hand corner of the computer monitor and started counting down when the experiment begun. The timer was shown throughout the experiment. Subjects were instructed that the experiment finished when timer value was "0.00" and their purpose was to earn time, reducing the timer and leading them to finish the experiment faster. Primary consequences were signaled by a message shown on the screen, and changes in the timer were implemented instantaneously. NSs (S+ and S-) were two colored disks (yellow and blue) with a diameter of 4 cm. Colors were counterbalanced across subjects. USs and CSs were shown above the central key (see Figure 1 for an illustration). Three squares with sides of 3 cm were used as response keys. Response keys were grey when they were inactive, and green when active. Two of the three keys were used as concurrent responses keys (during phases 1 and 3). A third response key was displayed above the two others, at the center of the screen (during phase 2). Throughout the experiment, the background of the interface was black.

2.3 Procedure

Instructions were written on paper and given to the participants after having obtained their informed consent. It was explained that their goal was to reduce a timer displayed on the computer screen by earning time. They were shown how to use the computer mouse to click on squares shown on the screen (responses keys) to earn time. It was specified that they had to make choices using the concurrent response keys, or just click on the central key during the other phase. Ratio values running on the response keys were specified. A schematic representation of the interface was given before participants began the experiment (see Appendix for exact instructions).

For all 50 participants, the experiment was composed of three phases. The first phase served to test the reinforcing value of time reduction by using a concurrent ratio schedule with one (side) key leading to time reduction and the other not. The second phase consisted of the pairing between NS and US. During this phase, participants were assigned to one of five groups (three experimental and two control groups, see below). Finally, the third phase was a test phase that consisted of a concurrent ratio schedule similar to the one used in phase 1, except that S+ and S- were delivered instead of the USs. Phase 3 was designed to assess the acquisition of reinforcing value by S+.

2.3.1 Phase 1 - Concurrent ratio schedule with time reduction as US

Phase 1 (illustrated in Figure 1, panel A) consisted of a concurrent fixed-ratio 12 schedule (FR12). Twelve clicks on one key led to a gain of 3 min (i.e. -3.00) and 12 clicks on the other key led to no gain (i.e. -0.00). The allocation of consequences was randomly determined when phase 1 started and remained fixed throughout this phase (i.e. a response key led to the same consequence throughout the entire phase). A choice began with an inter-trial interval (ITI) of 3 s during which the concurrent response keys were grey. After the ITI, the two keys became active, lighting up in green, and the completion of the FR12 on one of the two keys led to the corresponding consequence (gain of 3 min or no gain). The consequence consisted of a message shown for 2 s signaling the primary stimulus delivered (-3.00 or -0.00), and, when the primary stimulus was -3.00, the change in timer appeared at message onset. The delay between the ratio completion and US was 0.1 s. After a total of 10 US were delivered (i.e. 10 ratio completions), phase 1 finished. The number of responses made on each key by the participant was recorded. During this and the following two phases, the cursor position was not reset after each click or US presentation.

Please insert FIGURE 1 here,

2.3.2 Phase 2 – Conditioning

NS and US pairings were conducted in phase 2 under five different conditions (B1, B2, B3, F, and N). Figure 2 summarizes all the trials (continuous and partial reinforcement) for all the pairing conditions. Blocks represent a pairing trial (e.g. US-S+ pairing), and the color indicates the stimuli paired and the pairing condition. Hatched blocks represent trials where the CSs were delivered alone (see below for details).

In B1 condition (for Backward 1), US-S+ and noUS-S- backward pairings were presented in separate blocks (see Panel A of Figure 2). Pairings were conducted by using an operant contingency summarized in Figure 1, panel B. For US-S+ pairings, the pairing started with an inter-trial interval

(ITI) of 15 s followed by activation of the central key. On this key, completion of an FR 12 schedule led to US presentation followed by S+. The time interval between last ratio response and US and between US and S+ was 0.1 s. The US was presented for 2s and the S+ for 0.5. The pairing was the same for noUS-S- except that these stimuli were presented instead of US and S+. The B1 condition was a "unit" condition because the sequences of US-S+ pairings alternated with the sequences of noUS-S- pairings. In the B1 condition, the experiment could start with either US-S+ pairings or noUS-S- pairings, but critically either of these was given in blocks. Also note that (in this and other conditions) the deterministic schedule used early during Continuous Reinforcement was progressively faded into Partial Reinforcement, and hence the blocks were interspersed with S+ or S- alone trials. This was done to reduce generalization decrement and extinction during phase three of testing.

Please insert FIGURE 2 here,

The B2 condition was similar to the B1 condition except that the order of pairings changed, as illustrated in Figure 2 panel B. B2 was an "alternated" condition where the blocks (US-S+ and noUS-S-) alternated after each new pairing. While in B1 S+ was followed most of the time by a new US-S+ pairing and S- by a new noUS-S- pairing, in B2 S+ was followed (after the ITI) by noUS and S- by the US. There are two reasons for using this alternating procedure. First, the development of S+ reinforcing value in B1 could be explained by the Signal Hypothesis (SH) by S+ signaling the US presented at the end of the next block (i.e. trace conditioning). Consequently, in B2 S+ would signal noUS and thus should not develop a reinforcing value if this interpretation based on SH is correct. Secondly, Delay Reduction Theory (DRT) assumes that the more a CS signals delay reduction in US presentation, the more it will develop a reinforcing value. In B2, S- is closer (in the forward direction) to the next US, so S- should develop a greater reinforcing value over S+. Again, in B2 continuous reinforcement was administered early on, and this progressively turned into a partial reinforcement schedule with interspersed presentations of S+ or S- alone.

In the B3 condition, noUS was not presented and S+ and S- were both correlated with the US. Pairings in B3 started with an ITI of 20 s. During the ITI, S- was presented after 7 s of the ITI had elapsed (≈ 1/3 of the ITI). The ITI was then followed by central response key activation where a FR12 schedule was in force. Ratio completion led to the US followed by S+ as in B1 and B2. Again, this pairing was designed to test the SH and DRT. Indeed, by introducing S- during the ITI without noUS, the two CSs should become equally correlated with the US. Moreover, if a trace conditioning interpretation is used to explain a conditioned reinforcing effect of S+ in B1, because S- is closer (in forward direction) to the US than S+, according to SH and DRT it should develop greater reinforcing value over S+ as in B2. The F and N conditions were similar to B1 except that in the F condition S+-US and S- -noUS pairings were designed (forward pairings condition), and in N condition there was neither US nor noUS presentation (neutral condition).

At the end of phase 2, US and S+ (as well as noUS and S-) were paired ten times except in condition B3, where S+ and S- were paired ten times with the US. Note that partial reinforcement was introduced to reduce the Pavlovian extinction effect in phase 3, and CSs were delivered nine times alone in B1, B2, B3, and F conditions. Figure 2 summarizes the continuous and partial reinforcement. Thus, Phase 2 was divided into three periods (continuous reinforcement followed by two partial reinforcement periods). In the continuous reinforcement period, four pairings were designed between the US and S+ and between S- and noUS. In the first partial reinforcement period (partial reinforcement 1), the stimuli were paired three times and the CSs were presented alone three times. The order in pairings and CS alone trials was randomized. Finally, in the second partial reinforcement period (partial reinforcement 2), the stimuli were paired three times and the CSs were presented six times alone.

2.3.3 Phase 3 - Concurrent ratio schedule with S+ and S-

Phase 3 consisted of a concurrent FR12 schedule, with one key leading to S+ and the other to S-. The USs were not delivered in phase 3, and the keys were always reversed compared to phase 1, so the key leading to -3.00 in phase 1 led to S- in phase 3. Contingencies were similar to those designed in phase 1, except that CSs were the consequences instead of time reduction. Delay between the last response in the FR12 and the CS was 0.1 s and the CS was presented for 0.5 s as in phase 2. Phase 3 ended after 10 consequences. The number of responses on the response keys leading to S+ and S- was recorded.

2.4 Results

In phase 1, higher responding was expected on the response key followed by time reduction of -3.00 (US) in comparison with the response key delivering no time reduction, or -0.00 (no US). Results are summarized in Figure 3, which depicts the average number of responses ($\acute{x}$) and standard deviation (σ) for each condition on the response key delivering -3.00 (dark grey) or -0.00 (light grey). A 5 (conditions) x 2 (keys) mixed analysis of variance (ANOVA) revealed no effect of conditions, $F\ (4,45) = 0.00$, $p < .05$, a significant effect of keys, $F(1,45) = 600$, $p < .05$, and no interaction, $F\ (4,45) = 0.20$, $p < .05$. This pattern of results suggests that time reduction was able to selectively reinforce the key that led to it, and that there were no group differences, as expected.

Please insert FIGURE 3 here,

During phase 3, response keys delivered S+ and S- without the primary reinforcer. Again, we measured the number of responses on the two keys for each condition. The results are summarized in Figure 4, which depicts the average number of responses ($\acute{x}$) and standard deviation (σ) for each condition on the response key delivering S+ (dark grey) or S- (light grey). A 5 (conditions) x 2 (keys) mixed analysis of variance (ANOVA) revealed no effect of conditions, $F\ (4,45) = 1.00$, $p < .05$, a significant effect of keys, $F(1,45) = 168.65$, $p < .05$, and a significant interaction, F (4,45) = 20.49, $p < .05$. Simple effects analyses show a significant effect of keys in groups B1, (F(1,9) = 97.68, $p < .05$), B2, (F(1,9) = 135, $p < .05$), B3, (F(1,9) = 49.85, $p < .05$), and F, F(1,9) = 121.26, $p < .05$), but not in N, (F(1,9) = 1.95, $p > .05$). Moreover, an analysis conducted on S+ for groups B1, B2, B3, and F show no significant difference between the conditions (F(3,36) = 1.3, $p > .05$). In conclusion, conditioned reinforcement effects were demonstrated in both backward and forward conditions, but not in the neutral condition, and no differences were observed between the backward groups and the forward group.

Please insert FIGURE 4 here,

3. Discussion

The first evidence of conditioned reinforcement was reported in the first half of the 20th century, notably in the studies conducted by Bugelski (1938) and Skinner (1938). Since then, many procedures have been used to study conditioned reinforcement, but there is no agreement concerning the definition and function of a conditioned reinforcer, and the subject is still debated in the literature (Shahan, 2010). Explanations of conditioned reinforcement can be broadly divided into three hypotheses: (1) the conditioned reinforcement hypothesis, or CRH (Dinsmoor, 1983; Skinner, 1938, 1953), (2) the delay reduction theory, or DRT (Fantino, 2008; Fantino et al., 1993; Preston & Fantino, 1991), and finally (3) the signal hypothesis, or SH (Bolles, 1975; Davison & Baum, 2006; Longstreth, 1971; Schuster, 1969; Shahan, 2010). This experiment was designed to test these explanations.

In order to do that, we conducted an experiment with human subjects in a conditioned reinforcement paradigm. In addition to using forward pairings, as is done in most experiments on conditioned reinforcement, we used a backward conditioning procedure for the pairings between the NSs and the USs. Backward conditioning was used because it allowed us to directly test the hypotheses cited above. Theoretically speaking, backward conditioning should not lead to conditioned responding because the CS is correlated with a no-US period, or US absence. However, a large body of data has shown excitatory backward conditioning in animals (Ayres et al., 1987; Burkhardt, 1980; Chang et al., 2003; Heth, 1976; Mahoney & Ayres, 1976; Spetch et al., 1981, 1982; Tait & Saladin, 1986) in first-order paradigms, or with human subjects in second-order paradigms (Arcediano et al., 2003), and these results are incompatible with traditional associative learning models (Mackintosh, 1975; Pearce & Hall, 1980; Rescorla & Wagner, 1972). We studied backward conditioning in our conditioned reinforcement experiment to directly test the SH. Moreover, we included several conditions (such as B2 and B3) that allowed us to test the DRT.

Evidence for conditioned reinforcement as observed in B1 condition, where sequences of US-S+ pairings were followed by sequences of noUS-S- pairings, argue against the standard SH. Indeed, in the B1 condition S+ is predictive of the ITI and not of the US, a condition supposed to produce inhibitory rather than excitatory learning. The prediction is that S+ should not have become a conditioned reinforcer in B1. However, one might argue that S+ developed conditioned reinforcement properties because of trace conditioning between S+ presented at trial n and the US presented at trial n+1. This interpretation was tested in conditions B2 and B3. Indeed, in condition B2, S+ was followed after the ITI by noUS, and, in condition B3, S+ and S- were correlated with the US but S- was closer (in forward direction) than S+ to the US. Here, based on SH and on a trace conditioning interpretations of B1, a preference for S- should have been observed; but the results clearly show the opposite, with a strong preference for the response key producing S+. Data from conditions B2 and B3 also argue against DRT. Indeed, DRT assumes that the more a CS is correlated with reduction time from US onset, the more it will develop reinforcing properties. Here, because S- was closer (in forward direction) to the US, a preference for that stimulus should have been observed as well. Our results are not consistent with this expectation. Finally, we should underline that no differences were found between backward conditions (B1, B2, and B3) and the forward (F) condition, unlike most of the data on backward conditioning showing a weaker effect of that pairing procedure in comparison with a forward one (see for example Heth & Rescorla, 1973; Mahoney & Ayres, 1976; Romaniuk & Williams, 2000).

Contrary to SH and DRT, these results support CRH and its assumption that a predictive relationship is not necessary for the development of conditioned reinforcement effects by a CS. Following this assumption, Donahoe (2014; also see Donahoe & Vegas, 2004) have stated that the determinant for excitatory conditioning is the temporal contiguity (or overlap) between the CS and the UR produced by the US. Even if this overlap was not directly tested in the present experiment, strong temporal contiguity between US and S+ (rather than S-) led to conditioned reinforcement in phase 3. Overall, these results follow the large body of data showing first-order excitatory backward conditioning in animals or second-order excitatory backward conditioning in humans (Arcediano et al., 2003; Ayres et al., 1987; Burkhardt, 1980; Chang et al., 2003; Heth, 1976; Mahoney & Ayres, 1976; Spetch et al., 1981, 1982; Tait & Saladin, 1986) and argue against traditional learning models (Mackintosh, 1975; Pearce & Hall, 1980; Rescorla & Wagner, 1972). These models (like the SH) assume that the CS has to be a non-redundant predictor of the US to produce a conditioned response (CR), a postulate incompatible with excitatory backward conditioning. Rather, these data support recent considerations on the necessary and sufficient conditions for learning. These considerations can be found in work by Donahoe and colleagues cited above (Donahoe & Palmer, 2004; Donahoe & Vegas, 2004; Donahoe, 2014), but also in the Temporal Coding Hypothesis, or TCH (Arcediano & Miller, 2002). As an alternative to traditional learning models, TCH makes the assumption that temporal contiguity is a necessary and sufficient condition for learning. Moreover, TCH assumes that learning not only leads to the formation of a “What” association but also to a “When” association, so temporal attributes of an association are also encoded, leading to the formation of temporal maps when temporal information is integrated. Maps from different situations can be superimposed (i.e., integrated) during test and conditioned responding to a CS or excitatory/inhibitory value of a CS is determined by the predictive value of a temporal contingency (see the results on second-order excitatory backward conditioning by Arcediano et al., 2003 for an illustration). Based on these assumptions, the results observed in this study can easily be explained by TCH. The hypothesis would assume that an associations between the US and S+, and between the noUS and S- were formed in phase 2 because of the temporal contiguity between the events. Then, during phase 3, the early portions of the FR12 should predict the US occurrence, as it precedes S+ presentation. In other words, responding for S+ would become predictive of US delivery and S- of noUS delivery because of superimposed temporal maps.

This interpretation of our results based on temporal integration is also supported by a recent study conducted by Thrailkill and Shahan (2014) in a conditioned reinforcement paradigm. They trained two groups of rats to associate a CS1 with the US, with the critical difference between the two groups being that one group received delay CS1-US conditioning, while the second experienced trace CS1-----US pairings. As expected, the authors found a greater conditioned reinforcement effect for CS1 in a delay than in a trace conditioning group. However, when a second-order conditioning session was added between CS1 and a novel stimulus (CS2), testing of CS2 showed a greater conditioned reinforcement effect in the trace conditioning group than in the delay conditioning group. This finding is consistent with the idea that animals integrated CS1----US and CS1-CS2 associations. These results were interpreted by Thrailkill and Shahan as suggesting that animals encoded the temporal relation between CS1 and the US for the development of conditioned reinforcement effects with CS2 (CS2 being predictive of the US in the trace group but not in the delay group), and arguing for a temporal integration of intervals (Arcediano & Miller, 2002).

Finally, changes in liking of the CS achieved through evaluative conditioning (i.e. associatively induced changes in the value of a stimulus) can also explain the preference for S+ over S- during phase 3. Evaluative conditioning has been observed after backward conditioning (De

Houwer, Thomas, & Baeyens, 2001). Following this, because our CSs were repeatedly paired with USs with different value, and because we used a choice situation in our test session, we may assume that the choice to allocate responses to produce S+ was due to evaluative conditioning. A good test of this hypothesis would be to assess the number of trials effect in test session. Indeed, because evaluative conditioning is supposed to be insensitive to extinction, preference should be the same in the test session, for different numbers of CS presentations in the test session. Future experiments shall evaluate the merits of these predictions.

In summary, excitatory backward conditioning was assessed in a conditioned reinforcement paradigm. This experiment was conducted with human subjects, consisted of five conditions, and conditioned reinforcement value was assessed in a choice paradigm. The results suggest conditioned reinforcement effects in all backward conditions similar to those observed in the forward control condition, but not in the neutral condition. This study is the first demonstration in human subjects of conditioned reinforcement after backward conditioning, and one of the first studies showing first-order excitatory backward conditioning in human subjects.

References

Arcediano, F., & Miller, R.R. (2002). Some constraints for models of timing: a temporal coding hypothesis perspective. *Learning and Motivation, 33*, 105-123.

Arcediano, F., Escobar, M., & Miller, R.R. (2003). Temporal integration and temporal backward associations in human and nonhuman subjects. *Learning and Behavior, 31*, 242-256.

Ayres, J.J.B., Haddad, C., & Albert, M. (1987). One-trial excitatory backward conditioning as assessed by conditioned suppression of licking in rats: Concurrent observations of lick suppression and defensive behaviors. *Animal Learning and Behavior, 15*, 212-217.

Bersh, P.J. (1951). The influence of two variables upon the establishment of a secondary reinforcer for operant responses. *Journal of Experimental Psychology*, *41*, 62-73.

Bolles, R.C. (1975). *Theory of Motivation* (2nd Ed). New York: Harper & Row.

Bugelski, R. (1938). Extinction with and without sub-goal reinforcement. *Journal of Comparative Psychology, 26*, 121-134.

Burkhardt, P.E. (1980). One-trial backward fear conditioning in rats as a function of US intensity. *Bulletin of Psychonomic Society, 15*, 9-11.

Chang, R., Blaisdell, A.P., & Miller, R.R. (2003). Backward conditioning: Mediation by the context. *Journal of Experimental Psychology: Animal Behavior Process, 29*, 171-183.

Cowles, J.T. (1937). Food tokens as incentives for learning by chimpanzees. *Comparative Psychology Monograph*, *14*, 1-96.

Davison, M., & Baum, W.M. (2006). Do conditional reinforcers count? *Journal of Experimental Analysis of Behavior, 86*, 269-283.

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative learning of likes and dislikes: a review of 25 years of research on human evaluative conditioning. *Psychological Bulletin*, *127*, 853-869.

Dinsmoor, J.A. (1983). Observing and conditioned reinforcement. *Behavioral and Brain Sciences*, *6*, 693–728.

Donahoe, J.W. (2014). Evocation of behavioral change by the reinforcer is the critical event in both the classical and operant procedures. *International Journal of Comparative Psychology*, *27*, 537-543.

Donahoe, J.W., & Palmer, D.C. (2004). *Learning and complex behavior*. Richmond, VA: Ledgetop Publishing.

Donahoe, J.W., & Vegas, R. (2004). Pavolovian conditioning: the CS-UR relation. *Journal of Experimental Psychology: Animal Behavior Process, 30*, 17-33.

Egger, M.D., & Miller, N.E. (1963). When is a reward reinforcing? An experimental study of the information hypothesis. *Journal of Comparative and Physiological Psychology, 56*, 132-137.

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

Fantino, E. (1977). Conditioned reinforcement: Choice and information, in: Honig, W.K., & Staddon, J.E.R. (Eds), *Handbook of Operant Behavior*. Englewood Cliffs, NJ: Prentice-Hall, pp. 313-339.

Fantino, E. (2008). Choice, conditioned reinforcement, and the prius effect. *The Behavior Analyst, 31*, 95-111.

Fantino, E., Preston, R.A., & Dunn, R. (1993). Delay reduction: current status. *Journal of Experimental Analysis of Behavior, 50*, 159-169.

Fantino, E., & Silberberg, A. (2010). Revisiting the role of bad news in maintaining human observing behavior. *Journal of Experimental Analysis of Behavior, 93*, 157-170.

Gallistel, C.R., Craig, A.R., & Shahan, T.A. (2014). Temporal contingency. *Behavioural Processes, 101*, 89-96.

Gallistel, C.R., & Gibbon, J. (2000). Time, rate, and conditioning. *Psychological Review, 107*, 289-344.

Hackenberg, T.D. (2009). Token reinforcement: A review and analysis. *Journal of Experimental Analysis of Behavior, 91*, 257–286.

Hall, R.V., Lund, D., & Jackson, D. (1968). Effects of teacher attention on study behavior. *Journal of Applied Behavior Analysis, 1*, 1-12.

Heth, C.D. (1976). Simultaneous and backward fear conditioning as a function of number of CS-US pairings. *Journal of Experimental Psychology: Animal Behavior Process, 2*, 117-129.

Heth, C.D., & Rescorla, R.A. (1973). Simultaneous and backward fear conditioning in the rat. *Journal of Comparative and Physiological Psychology, 82*, 434-443.

Hull, C.L. (1943). *Principles of Behavior*. Appleton Century-Crofts. New-York.

Jenkins, W.O. (1950). A temporal gradient of derived reinforcement. *American Journal of Psychology, 63*, 237-243.

Jones, K.M., Drew, H.A., & Weber, N.L. (2000). Noncontingent peer attention as treatment for disruptive classroom behavior. *Journal of Applied Behavior Analysis, 33*, 343-346.

Kamin, L.J. (1969). Predictability, attention, surprise, and conditioning. In B.A. Campbell & R.M. Church (Eds.), *Punishment and aversive behavior* (pp. 279-296). New York: Appleton-Century-Crofts.

Kazdin, A.E. (1977). *The Token Economy*. New-York: Plenum Press.

Kelleher, R.T., & Gollub, L.R. (1962). A review of positive conditioned reinforcement. *Journal of Experimental Analysis of Behavior, 6*, 543-597.

Leung, J.P. (1989). Psychological distance to reward: a human replication. *Journal of Experimental Analysis of Behavior, 51*, 343-352.

Leung, J.P. (1993). Psychological distance to reward: segmentation of aperiodic schedules of reinforcement. *Journal of Experimental Analysis of Behavior, 59*, 401-410.

Levin, S.M., & Sterner, K. (1966). The acquisition of positive secondary reinforcement in human subjects. *Psychonomic Science, 6*, 47-48.

Long, E.R. (1963). Chained and tamdem scheduling in children. *Journal of Experimental Analysis of Behavior, 6*, 459-472.

Longstreth, L.E. (1971). A cognitive interpretation of secondary reinforcement. In J.K. Cole (Ed), *Nebraska Symposium on Motivation, Vol 19* (pp. 33-81). Lincoln, NE: University of Nebraska Press.

Mackintosh, N.J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcements. *Psychological Review, 82*, 276-298.

Mahoney, W.J., & Ayres, J.J.B. (1976). One-trial simultaneous and backward fear conditioning as reflected in conditioned suppression of licking in rats. *Animal Learning and Behavior, 4*, 351-362.

Myers, N.A., & Myers, J.L. (1962). Effects of secondary reinforcement schedules in extinction on children's responding. *Journal of Experimental Psychology, 64,* 586-588.

Northup, J.K., Broussard, C., Jones, K., & Herring, M. (1995). The differential effects of teacher and peer attention on the disruptive classroom behavior of three children with a diagnosis of attention deficit hyperactivity disorder. *Journal of Applied Behavior Analysis, 28*, 227-228.

Pearce, J.M., & Hall, G. (1980). A model of Pavlovian learning: Variations in the effectiveness of conditioned but not unconditioned stimuli. *Psychological Review, 87,* 532-552.

Preston, R.A., & Fantino, E. (1991). Conditioned reinforcement value and choice. *Journal of Experimental Psychology, 55,* 155-175.

Rescorla, R.A., & Wagner, A.R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonrienforcement. In A.H. Black & W.F. Prokasy (Eds). *Classical conditioning II: Current research and theory* (pp. 64-69). New York: Appleton-Century-Crofts.

Romaniuk, C.B., & Williams, D.A. (2000). Conditioning across the duration of a backward conditioned stimulus. *Journal of Experimental Psychology: Animal Behavior Processes, 26*, 454-461.

Savastano, H.I., & Miller, R.R. (1998). Time as content in Pavlovian conditioning. *Behavioural Processes, 44*, 147-162.

Schuster, R.H. (1969). A functional analysis of conditioned reinforcement, in: Hendry, D.P. (Eds), *Conditioned Reinforcement*. Homewood, IL: The Dorsey Press, pp. 192-235.

Shahan, T.A. (2010). Conditioned reinforcement and response strength. *Journal of Experimental Analysis of Behavior, 93,* 269-289.

Shahan, T.A., & Podlesnik, C.A. (2005). Rate of conditioned reinforcement affects observing rate but not resistance to change. *Journal of Experimental Analysis of Behavior, 84*, 1–17.

Shahan, T.A., & Podlesnik, C.A. (2008). Conditioned reinforcement value and resistance to change. *Journal of Experimental Analysis of Behavior, 89*, 263–298.

Skinner, B.F. (1938). *The Behavior of Organisms*. New-York: Appleton-Century-Crofts.

Skinner, B.F. (1953). *Science and Human Behavior*. New-York: Free Press.

Sosa, R., dos Santos, C.V., & Flores, C. (2011). Training a new response using conditioned reinforcement. *Behavioural Processes, 87*, 231-236.

Spence, K.W. (1950). Cognitive versus stimulus-response theories of learning. *Psychological Review, 57*, 159-172.

Spetch, M.L., Terlecki, L.J., Pinel, J.P.J., Wilkie, D.M., & Treit, D. (1982). Excitatory backward conditioning of defensive burying in rats. *Bulletin of Psychonomic Society, 19*, 111–114.

Spetch, M.L., Wilkie, D.M., & Pinel, J.P.J. (1981). Backward conditioning: A reevaluation of the empirical evidence. *Psychological Bulletin, 89*, 163-175.

Tait, R.W., & Saladin, M.E. (1986). Concurrent development of excitatory and inhibitory associations during backward conditioning. *Animal Learning and Behavior, 14*, 133–137.

Thrailkill, E.A., & Shahan, T.A. (2014). Temporal integration and instrumental conditioned reinforcement. *Learning and Behavior, 42*, 201-208.

Tolman, E.C., & Brunswik, E. (1935). The organism and the causal texture of the environment. *Psychological Review, 42*, 43-77.

Urushihara, K. (2004). Excitatory backward conditioning in an appetitive conditioned reinforcement preparation with rats. *Behavioural Processes, 67*, 477-489.

Williams, B.A. (1994). Conditioned reinforcement: experimental and theoretical issues. *The Behavior Analyst, 17,* 261-285.

Wolfe, J.B. (1936). Effectiveness of token rewards for chimpanzees. *Comparative Psychology Monographs, 12*, 1-72.

Appendix

Instructions (*translated from French to English*).

"Thank you for your participation in this study on human behavior.

When the experiment will start, on the computer screen you will see a timer with the value: "180.00" (time in minute). The experiment will finish when the timer value will be "0.00". Your aim in this experiment will be to earn at several occasions a specific quantity of time (-03.00) that will reduce the timer value. By reducing this value, the experiment will finish faster.

During the experiment, depending on your actions, you may earn time (-03.00) or not (-00.00). On the computer screen, 3 squares will be shown (see the illustration below). Squares could be active (green color) or inactive (grey color). When the two bottom squares are active, you will have to make a choice by clicking 12 times with the computer mouse on one of them. Otherwise, only the central square will be active.

At different moments, two disks of different colors could be shown. It is not requiring for you to click on that disks. Finally, when you earn time or not, a message is shown on the screen and change in the timer is implemented at the moment."

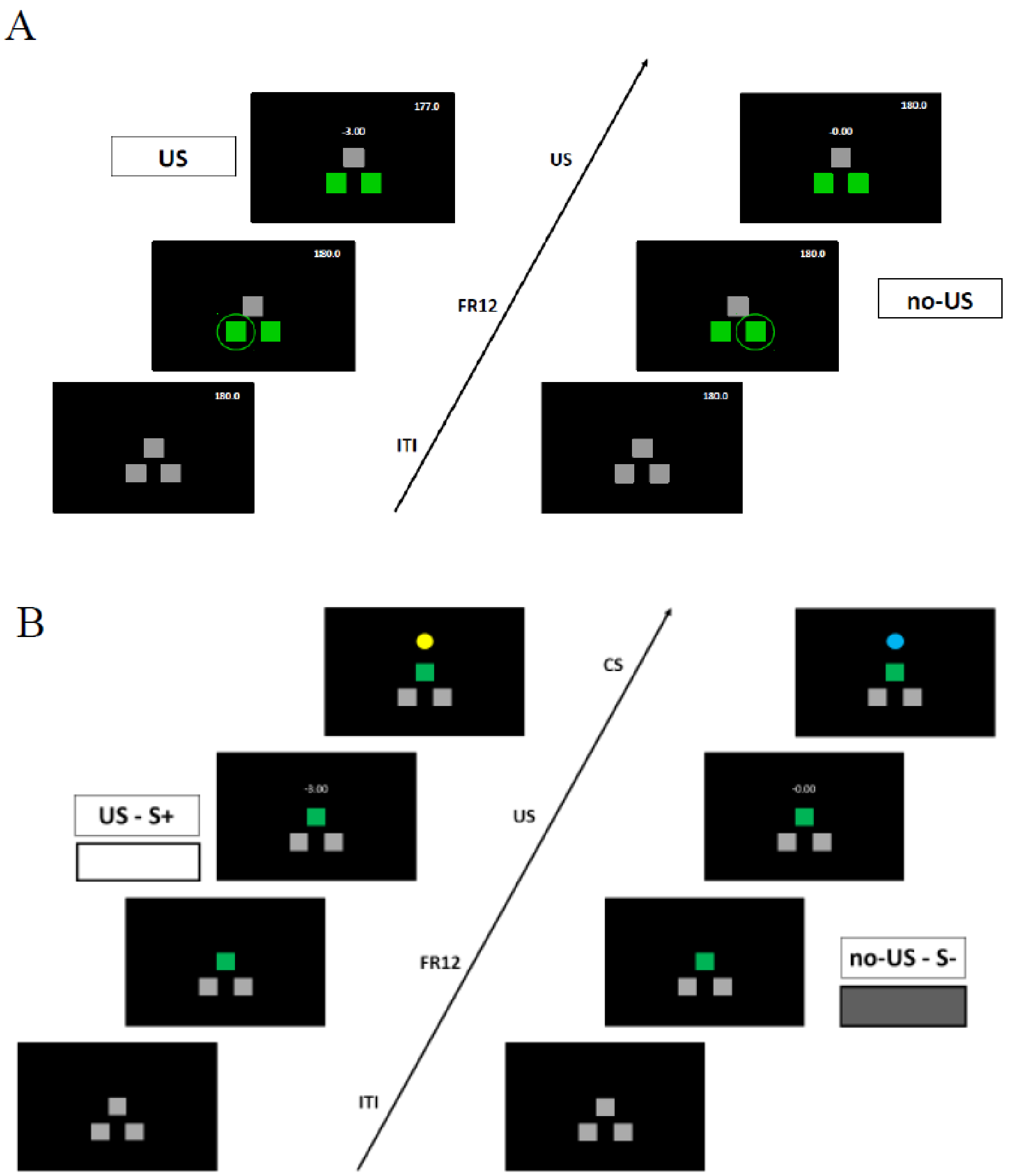
Figure Captions

Figure 1 – Schematic representation of the concurrent ratio schedule with time reduction as the US (Panel A), and of the US-S+ and noUS-S- pairings used in the B1 and B2 conditions (Panel B).

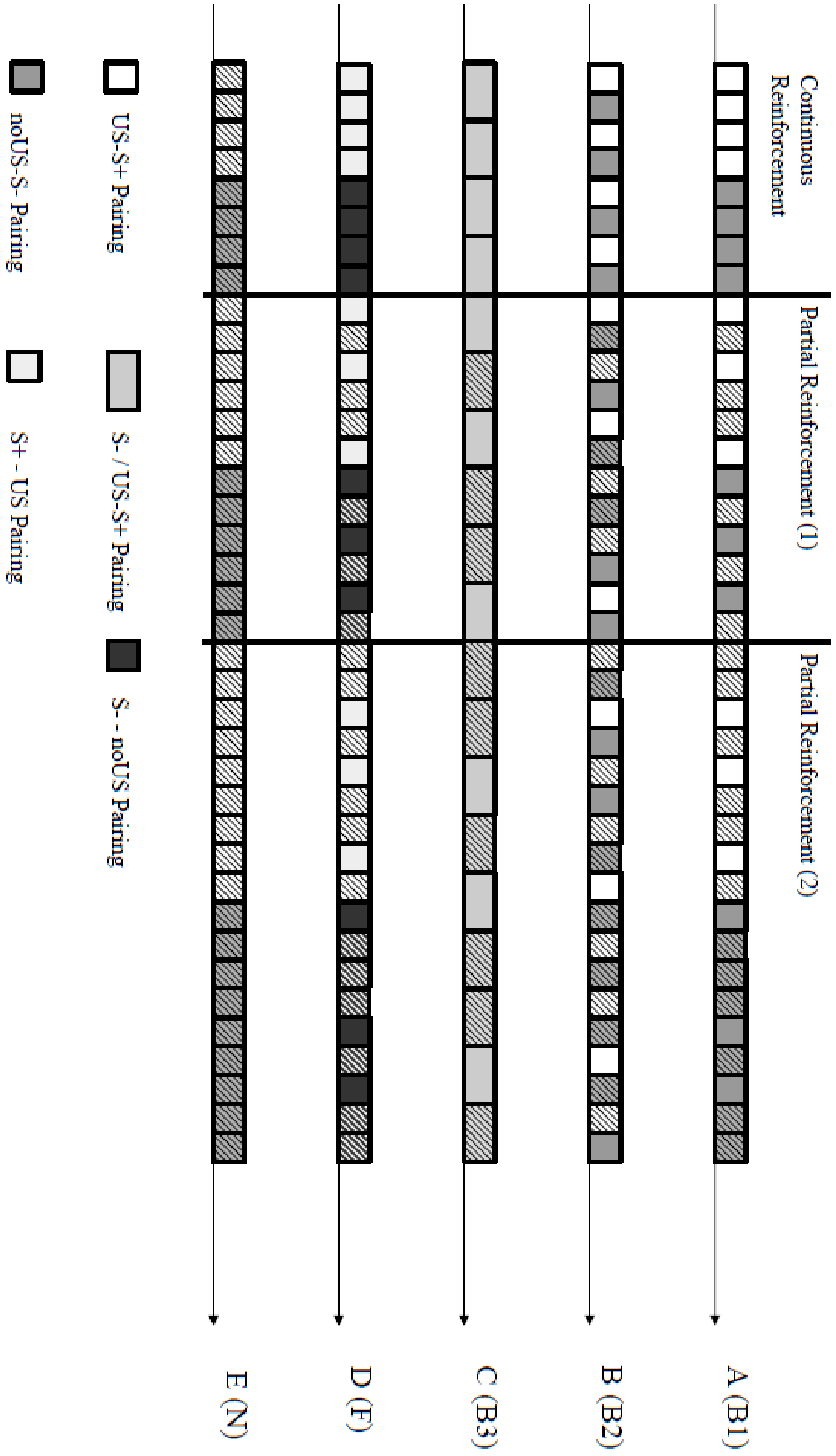
Figure 2 - Schematic representation of the sequences designed for the CS and the US pairings, for each condition. Blocks represent pairing trials and hatched blocks represent trials where a CS was presented alone.

Figure 3 - Results observed during phase 1 of training show the average number of responses ($\acute{x}$) and standard deviation (σ), for each condition, on the response key leading to -3.00 or -0.00.

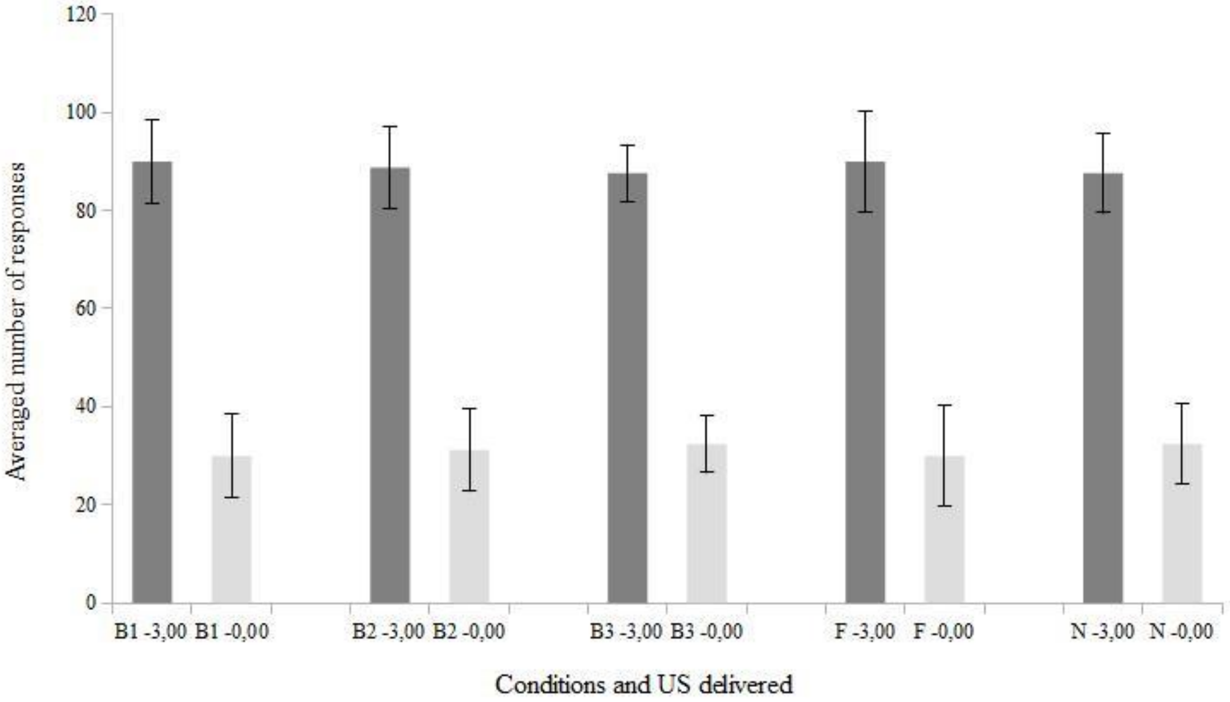
Figure 4 - Results observed during phase 3 show the average number of responses ($\acute{x}$) and standard deviation (σ), for each condition, on the response key leading to S+ or S-.



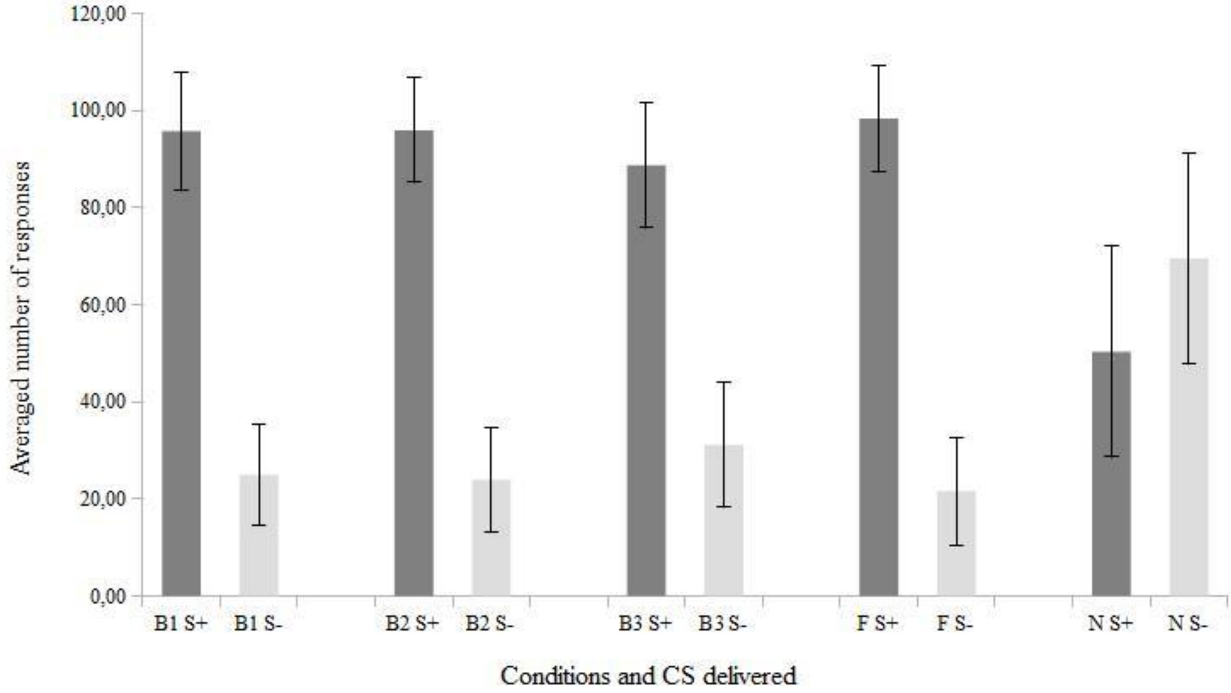
Prével, Rivière, Darcheville, & Urcelay - Figure 1



Prével, Rivière, Darcheville, & Urcelay - Figure 2



Prével, Rivière, Darcheville, & Urcelay - Figure 3



Prével, Rivière, Darcheville, & Urcelay - Figure 4