

THE COMBINATION OF APPETITIVE AND AVERSIVE REINFORCERS AND THE NATURE OF THEIR INTERACTION DURING AUDITORY LEARNING

A. ILANGO,^a W. WETZEL,^a H. SCHEICH^a AND F. W. OHL^{a,b*}

^aLeibniz Institute for Neurobiology, Brennekestr. 6, D-39118 Magdeburg, Germany

^bInstitute of Biology, Otto-von-Guericke University, Magdeburg, Germany

Abstract—Learned changes in behavior can be elicited by either appetitive or aversive reinforcers. It is, however, not clear whether the two types of motivation, (approaching appetitive stimuli and avoiding aversive stimuli) drive learning in the same or different ways, nor is their interaction understood in situations where the two types are combined in a single experiment. To investigate this question we have developed a novel learning paradigm for Mongolian gerbils, which not only allows rewards and punishments to be presented in isolation or in combination with each other, but also can use these opposite reinforcers to drive the same learned behavior. Specifically, we studied learning of tone-conditioned hurdle crossing in a shuttle box driven by either an appetitive reinforcer (brain stimulation reward) or an aversive reinforcer (electrical footshock), or by a combination of both. Combination of the two reinforcers potentiated speed of acquisition, led to maximum possible performance, and delayed extinction as compared to either reinforcer alone. Additional experiments, using partial reinforcement protocols and experiments in which one of the reinforcers was omitted after the animals had been previously trained with the combination of both reinforcers, indicated that appetitive and aversive reinforcers operated together but acted in different ways: in this particular experimental context, punishment appeared to be more effective for initial acquisition and reward more effective to maintain a high level of conditioned responses (CRs). The results imply that learning mechanisms in problem solving were maximally effective when the initial punishment of mistakes was combined with the subsequent rewarding of correct performance. © 2010 IBRO. Published by Elsevier Ltd. All rights reserved.

Key words: reinforcement learning, rodent, reward, punishment, shuttle-box.

Conditioning involves the association of neutral stimuli with appetitive or aversive reinforcers. Animals direct their behaviors, in both natural and laboratory situations (e.g. instrumental conditioning experiments), in such a way as to

*Correspondence to: F. W. Ohl, Leibniz Institute for Neurobiology, Brennekestr. 6, D-39118 Magdeburg, Germany. Tel: +49-391-6263-329; fax: +49-391-6263-328.

E-mail address: frank.ohl@ifn-magdeburg.de (F. W. Ohl).

Abbreviations: BSR, brain stimulation reward; CR, conditioned response; CS, conditioned stimulus; FS, foot shock; RT, reaction time; VTA, ventral tegmental area.

0306-4522/10 \$ - see front matter © 2010 IBRO. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuroscience.2010.01.010

obtain appetitive reinforcers (“rewards”) and avoid aversive reinforcers (“punishments”). In most animal conditioning experiments, behavioral measures of conditioning and of brain systems have been studied with one type of reinforcer (appetitive or aversive) only. Hence, the nature of the interaction between appetitive and aversive reinforcers during associative learning in the same experimental situation is not well understood. Scrutinizing this interaction experimentally meets with substantial difficulties (see Dickinson, 1976; Mackintosh, 1983; Magoon and Critchfield, 2008 for an overview of the underlying theoretical problems). On the procedural side there has been a lack of learning paradigms that train the same behavior using both appetitive and aversive reinforcers delivered with the same temporal contingency and titrated to achieve comparable effects, such that their combinatorial influence can be quantified. Consequently, most classical work on the subject has relied on indirect methods, typically utilizing sequential interaction between reward-driven and punishment-driven tasks.

Early work by Konorski and collaborators on stimulus approach and withdrawal proposed that the interaction between appetitive and aversive reinforcers is mutually inhibitory in nature (Konorski and Szwejkowska, 1956; Konorski, 1967). Subsequent studies addressed the behavioral influence of stimuli associated with one type of reinforcer on stimuli associated with the other, using summation, retardation and counter-conditioning procedures (Dickinson and Pearce, 1977; Dickinson and Dearing, 1979; Mackintosh, 1983). Here also, an aversive stimulus was observed to suppress an appetitive response, and an appetitive stimulus was observed to suppress an aversive response (Estes and Skinner, 1941; Dickinson and Pearce, 1977).

Scavio (1974) demonstrated that preconditioning stimuli with shock impaired the further development of an appetitive response in rabbits. Appetitive conditioning and avoidance learning using the same stimuli indicated that appetitive and aversive conditioned motivational states interact subtractively (Bull, 1970). A conditioned stimulus (CS) associated with shock can inhibit the association of a CS paired with the omission of expected food reward (Dickinson and Dearing, 1979). The underlying motivational states of reinforcer interaction were also investigated using stimulus preexposure or preconditioning with one reinforcer. Fear conditioning was greatly enhanced if the CS was previously paired with food (Dickinson, 1977).

The above studies strengthen the argument that appetitive and aversive reinforcers can indeed interact. How-

ever, once the behavior is learned, intrinsic motivation also drives and strengthens associative learning (cf. Rolls, 2008). It has been demonstrated that, irrespective of reinforcer presentation, response-contingent neutral stimuli also can have intrinsic reinforcing properties (Reed et al., 1996). Stimulus generalization gradients in appetitive and aversive reinforcement investigated with two different responses showed that gradients of effect were different for the two types of reinforcers (Hearst, 1960).

In order to compare the potentially different roles of reward and punishment, it is desirable to develop behavioral procedures which can incorporate both types of reinforcers within the same training session (e.g. Magoon and Critchfield, 2008; Morrison and Salzman, 2009) as most previous experiments measured the effect of one reinforcer on the previously established conditioned response (CR) by the other reinforcer. Previous work had studied excitatory or inhibitory interactions between sequential reward- and punishment-driven learning processes (Dickinson, 1976, 1977; Dickinson and Mackintosh, 1978), concurrent schedules of reward and punishment without conditioned stimuli (Kelleher and Cook, 1959; Olds and Olds, 1962), combinations with secondary reinforcers associated with the opposite valence (Morris, 1975; Baron et al., 1977), and non-contingent schedules of aversive and appetitive reinforcers (Stein, 1965; Margules and Stein, 1968; Carder, 1970; Castro-Alamancos and Borrell, 1992). The design of the present set of experiments allowed us to demonstrate an equivalence of reward and relief from punishment in the sense of two forces acting towards a convergent effect. At the same time we could dissociate their different contributions to the early and late phases of a learning process.

Conventional reinforcers such as food and footshock (FS) involve different behavioral contingencies which are not easily combined in the same experiment and involve different forms of information processing. Also, a principal difference between appetitive and aversive reinforcers is that the effect of appetitive reinforcers typically saturates with prolonged presentation while the effect of aversive reinforcers does not. If we assume that avoidance responses must come under the control of additional positive incentives, the concept of combining “carrot and stick” arises: how effective is learning if an aversive stimulus is experienced upon unsuccessful avoidance and an appetitive stimulus is experienced upon successful avoidance?

Efforts have already been made to demonstrate the facilitation of learning using non-contingent application of positively reinforcing brain stimulation on aversive avoidance behavior (Margules and Stein, 1968; Castro-Alamancos and Borrell, 1992). Also, rewarding brain stimulation reduced the aversive reinforcing property of electric shock when both reinforcers were paired (Cox and Valenstein, 1965; Carr and Coons, 1982). Positively reinforcing posterior hypothalamic brain stimulation was used as a CS to signal the aversive FS; however, it further reduced the self stimulation performance during the post-conditioning sessions (Mogenson and Morrison, 1962). Facilitation of shuttle-box avoidance learning was observed if the animals

were allowed to self-stimulate for lateral hypothalamic brain stimulation immediately after the avoidance learning sessions (Aldavert-Vera et al., 1997). The difference between earlier studies and our investigation is the following: we deliver the internal reward for successful avoidance in a similar contingent way as the FS for unsuccessful avoidance and thereby address the nature of their interaction.

Here, we addressed the interaction of appetitive and aversive reinforcers in Mongolian gerbils (*Meriones unguiculatus*) during 2-way hurdle crossing in a shuttle box using auditory stimuli as CS, FS as an aversive reinforcer and brain stimulation reward (BSR), that is electrical stimulation of the ventral tegmental area (VTA), as an appetitive reinforcer. Our previous studies have demonstrated that Mongolian gerbils are suitable animal model to investigate auditory learning (Wetzel et al., 1998, 2008; Ohl et al., 1999, 2001). In preparatory experiments, current strengths for FS and VTA stimulation were separately calibrated to produce the same asymptotic level of behavioral performance in individual animals. After studying the effect of both reinforcers separately, we addressed reinforcer integration using FS as an aversive reinforcer upon each unsuccessful trial and BSR as an appetitive reinforcer upon each successful trial. This matched power of the reinforcers, in principle, allowed us to determine any possible type of interaction in combined experiments. FS-reinforced learning is initially dominated by aversive experience which leads to subsequent relief upon successful avoidance. Therefore, the primary question we addressed was whether effects from punishment and rewards inhibit each other, or alternatively, whether the relief from punishment and receipt of reward input facilitate the learning process (equivalence hypothesis) (Dinsmoor, 2001).

The dopamine signal from the midbrain substantia nigra pars compacta (SNc) and the VTA contributes to associative learning processes in which the exact timing of the reinforcement is vital (Schultz et al., 1997). Consistent with the role of dopamine as an encoder of stimulus-response associations which drive an instrumental act, dopamine medications in neuropsychiatric patients improved learning from success (Frank et al., 2004).

To obtain a comprehensive understanding of the effects of combining appetitive and aversive reinforcers to drive the associative learning we studied (1) acquisition and extinction using separate and combined reinforcers, (2) omission of one reinforcer in the combination experiment after animals had reached maximum performance followed by omission of the remaining reinforcer (extinction), and (3) omissions after partial reinforcement procedures.

EXPERIMENTAL PROCEDURES

Animals

A total of 80 adult male Mongolian gerbils (*Meriones unguiculatus*) obtained from Tumblebrook Farms, West Brookfield, MA, USA (age: 3–6 months, weight: 85–105 g) were used in this study. Gerbils were individually housed 3 days before experiments started and maintained on a 12 h light/dark cycle (light on 7:00–19:00 h) throughout the experiment. All experimental procedures

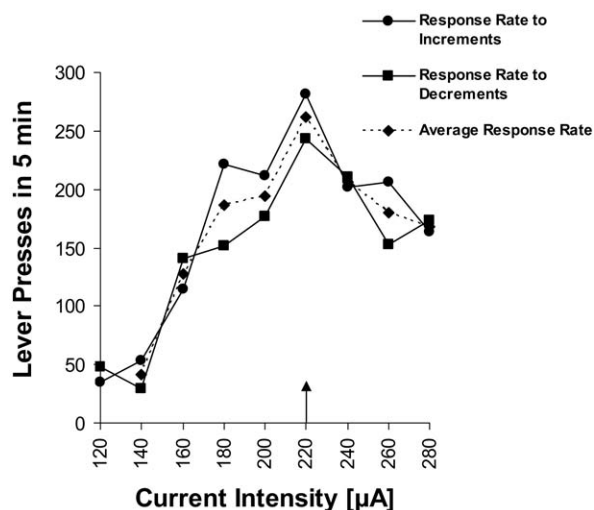


Fig. 1. Exemplary illustration of the procedure preceding shuttle-box training for determining the stimulation current strength for the brain stimulation reward. Plotted is the lever pressing rate of one gerbil against stimulation current intensity used for VTA stimulation. Stimulation current intensity was systematically increased from low levels to high levels (circles) and systematically decreased from high levels to low levels (squares). Diamond symbols indicate the mean rates of lever pressing of the increasing and decreasing series of measurements. The current intensity that produced maximal lever pressing rate (arrow) was used in the later shuttle-box experiments.

were approved by the Ethics Committee of the State of Sachsen-Anhalt, Germany.

Surgical procedures

Surgery and implantation of electrodes were performed under ketamine (100 mg/kg) (Ratiopharm GmbH, Ulm, Germany) and xylazine (5 mg/kg) (Bayer Vital GmbH, Leverkusen, Germany) anesthesia. Animals were fixed in a stereotaxic frame (David Kopf Instruments, USA). Bipolar stimulation electrodes with the tips separated by ~0.2 mm were custom made from Teflon-insulated stainless steel microwires (diameter: 140 µm; Science Products GmbH, Germany) and implanted at the level of the VTA (2.6 mm posterior, 1.3 mm lateral, 5.0 mm ventral to bregma) with the incisor bar set at -5, according to the stereotaxic atlas for gerbil by Loskota et al. (1974). The electrode was fixed in place with dental acrylic cement, and the procedure was repeated in the

opposite hemisphere. The first self-stimulation training session followed after a 4-day recovery period.

Intracranial self-stimulation

Self-stimulation training was conducted in a custom made operant chamber (18×18×23 cm) with a metal lever in the right lower corner. Gerbils were trained to press the lever for a brief VTA stimulation (200 ms train; 20 biphasic pulses of 0.2 ms duration at 100 Hz). Within 2–5 days, the animal learned to press the lever for VTA self-stimulation. Electrodes in both hemispheres were assessed initially. The electrode which supported maximum lever pressing performance in the absence of side effects was chosen for further optimization and shuttle-box learning.

After the gerbils showed lever pressing at stable rates, the optimum stimulus intensity was determined for each individual using ascending and descending stimulation intensities (step size 20 µA; each duration used for 5 min; Fig. 1) in two separate daily sessions. The current intensity which showed the maximum lever press rate (mean response rate between the ascending and descending stimulus intensities) was selected as the optimum intensity. The final session consisted of training the subjects with that optimum intensity for 30 min duration. This optimal intensity was used throughout the following learning tasks in the shuttle-box. Animals that did not reach the criterion (900 bar presses in 30 min within five consecutive days) were excluded from the study. After at least 3 days of rest, the intracranial self-stimulation (ICSS) responders ($n=54$) were allotted to various shuttle-box learning groups. The current strength determined in this way was found, when used as appetitive reinforcer in the conditioning experiments, to produce a similar level of final performance as in the FS-conditioned animals.

Shuttle box learning

Identical behavior, that is, tone-conditioned hurdle crossing driven by appetitive, aversive or the combination of both reinforcers was investigated using shuttle-box conditioning procedures. After that, the nature of their interaction was investigated by omission experiments (Table 1 for overview). Animals were trained in a shuttle-box (38×19×22.5 cm) (Hasomed GmbH, Magdeburg, Germany) which had two compartments separated by a 6 cm high hurdle. Each daily session consisted of 60 trials with an intertrial interval duration of 20–24 s. A session began with a 3 min habituation period. The CS was a series of 2 kHz pure tones (6 s, 200 ms tone duration, 300 ms inter-tone interval). When FS was used as a reinforcer, a FS of maximally 4 s was applied through the grid floor at the end of the CS in case the animal did not cross the hurdle during the 6 s CS presentation. The FS was switched off when the animal escaped, that is crossed the hurdle during FS presentation.

Table 1. Experimental scheme for the three main experiments

Experimental group	Associated reinforcers		
	Session 1–8	Session 9–13	Session 14–18
Experiment 1			
Group 1 (FS)	FS	Extinction	—
Group 2 (BSR)	BSR	Extinction	—
Group 3 (FS+BSR)	FS+BSR	Extinction	—
Experiment 2			
Group 4 (FS+BSR→BSR)	FS+BSR	BSR (omission of FS)	Extinction
Group 5 (FS+BSR→FS)	FS+BSR	FS (omission of BSR)	Extinction
Experiment 3			
Partial reinforcement			
Group 6 (FS+BSR→BSR)	FS+BSR	BSR (omission of FS)	Extinction
Group 7 (FS+BSR→FS)	FS+BSR	FS (omission of BSR)	Extinction

The intensity of FS was slowly raised from 400–600 μA during the first training session. Since the rate of conditioning also depends on the change in unconditioned stimulus intensity, we subsequently maintained a constant FS intensity of 600 μA . The appetitive reinforcer consisted of five blocks of 200 ms trains of electric pulses for brain stimulation reward with inter-stimulus interval of 300 ms. This pulse train was automatically delivered without delay when the animal reached the other shuttle-box compartment by tone-conditioned hurdle crossing. In different groups, either the appetitive reinforcer, the aversive reinforcer, or the combination of both reinforcers was applied. A flexible cable connected with a swivel allowed the electrical brain stimulation and easy movement during shuttle-box learning. The electrical stimulation was delivered by an isolated pulse stimulator (Model 2100, A-M Systems Inc, Carlsborg, WA, USA). Crossing the hurdle during the CS presentation was considered as a CR. The CR rates (number of CR/number of trials) and mean response latencies (times of reaching the new compartment after hurdle crossing) were analyzed in each session. In the first group, the CS was followed by the aversive FS if the animal did not cross the hurdle within 6 s. In the second group, a response was considered a CR if a hurdle crossing occurred within 6 s of the CS period, in which case, the CR was immediately followed by BSR. In the third group, trained with both FS and BSR, each successful CR was followed by BSR and each failure to produce a CR was punished by FS. So, they received the reward for successful avoidance. The three groups mentioned above (FS: footshock alone, BSR: brain stimulation reward alone, FS+BSR: footshock and brain stimulation reward combined), were trained in eight acquisition sessions followed by five extinction sessions in which reinforcers were not associated.

The fourth and fifth groups were trained with both reinforcers like the FS+BSR group. After they had been trained to reach maximum performance, in one group we omitted the BSR and in the other group we omitted the FS during sessions 9–13. This was followed by extinction training (sessions 14–18) in which we removed the remaining reinforcer, presenting only the CS.

The abovementioned experiments were repeated under partial reinforcement conditions. To determine the optimal probability of combined reinforcer presentation for the partial conditioning procedure, pilot experiments were first conducted in non-operated animals studying the effect of probability of FS on the CR rate (Fig. 4). Four groups of gerbils were trained with FS occurrence probabilities of 0.1, 0.15, 0.2 and 0.5, respectively. Since, in these experiments, FS probability of 0.15 produced asymptotic CR rates between 30% and 50%, and FS probability of 0.2 produced asymptotic CR rates between 50% and 70% (Fig. 4), we decided to use a FS probability of 0.15 for training with the combined appetitive-aversive reinforcer. This choice accounted for the expected increase in asymptotic CR level when FS motivation would be complemented by BSR to achieve combined reinforcement. Five (out of 15) animals did not learn the task and were excluded from the study. The sixth and seventh groups of animals were trained with BSR for hits and FS for misses (FS+BSR) with the partial reinforcement schedule, that is, in nine out of 60 trials ($P=0.15$), the hits were rewarded and the misses were punished according to a pseudorandom schedule. After reaching the desired level, like in the other groups, one reinforcer was omitted during sessions 9–13. This was followed by extinction sessions (14–18) in which the remaining reinforcer was also removed.

Data analysis

CR rates and response latencies were evaluated. To illustrate the learning curves, mean CR rate over daily sessions in each group, during acquisition, omission of one reinforcer and extinction was

plotted. Additionally, we measured the hurdle crossing rate during the intertrial interval. The behavioral data from all experiments were analyzed using repeated measures analysis of variance (ANOVA) with two within-subject factors (GROUP and SESSION), using SPSS software for windows (version 8.0). The degrees of freedom was corrected to more conservative values using the Greenhouse-Geiser correction procedure. After confirmation of significant main effect, Tukey post hoc tests were performed to assess pairwise differences among groups.

Histology

At the conclusion of behavioral experiments, the gerbils were deeply anesthetized with ketamine and xylazine mixture, and 200 μA current was injected for 15 min through the brain stimulation electrode. Thereafter, the animals were decapitated. The brains were rapidly isolated and frozen in isopentane immersed in liquid nitrogen and finally stored at $-20\text{ }^{\circ}\text{C}$. Subsequently, coronal sections of 40 μm roughly posterior to bregma were made using a sliding microtome (Leica CM 3050S, Leica Microsystems Nussloch GmbH, Karlsruhe, Germany). Nissl staining combined with Prussian blue was performed to reveal the ion deposits around the electrode tips (Fig. 6). The locations of the electrode tips were determined with reference to the stereotaxic atlas of gerbils (Lorskota et al., 1974).

RESULTS

All results were obtained in an auditory shuttle-box learning paradigm. In a 2-compartment shuttle-box, Mongolian gerbils were trained to change the current compartment by crossing a hurdle as the CR to the onset of a series of pure tones. Three experiments were conducted to investigate potentially different effects of appetitive and aversive reinforcers as well as the nature of their interaction when both types of reinforcers were combined within a single session. In Experiment 1, acquisition and extinction of the CR were studied in three experimental groups, using either reward of hits by electrical BSR, punishment of misses by electrical footshock (FS), or a combination of both in the same session (FS+BSR). In Experiment 2, effects of the omission of one type of reinforcer (appetitive or aversive) in animals that had been trained using the combined appetitive-aversive reinforcers were studied using a continuous reinforcement schedule. In Experiment 3, the analogous omission experiment was conducted under a partial reinforcement schedule. Table 1 gives an overview of the experimental groups. In all experiments, CR rates and reaction times were analyzed using a GROUP \times SESSION repeated measures analysis of variance (ANOVA) with SESSION as the repeated factor.

Experiment 1. Effects of appetitive, aversive, or combined appetitive-aversive reinforcers on acquisition and extinction of the CR

In three experimental groups, animals were trained by appetitive reinforcer (BSR on hits), aversive reinforcer (FS on misses), or combination of both reinforcer (BSR on hits and FS on misses) in the tone-cued shuttle-box task. Fig. 2a shows the mean CR rates during acquisition and extinction of the CR.

Analysis of the CR rates during the acquisition sessions (1–8) indicated a significant GROUP \times SESSION

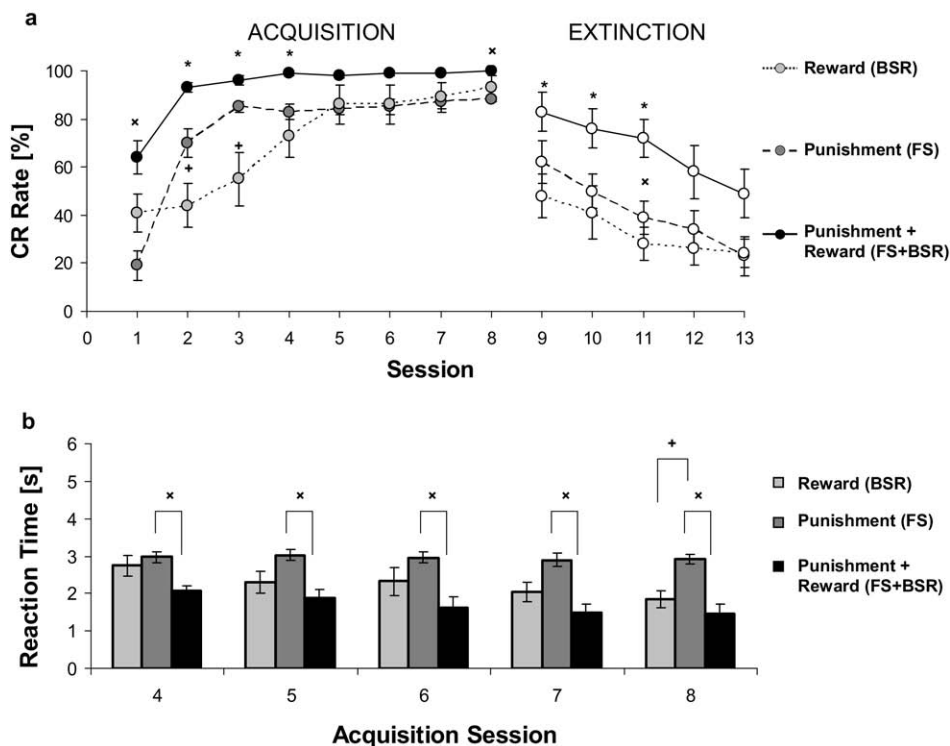


Fig. 2. Effect of reinforcer type on acquisition and extinction of a CR. In the three groups ($n=6$ in each group), learning was driven by FS punishment following misses (FS, dark grey), brain stimulation reward following hits (BSR, light grey), or both (FS+BSR, black). (a) Mean CR rates (\pm SEM) plotted against training session. The break in the graphs indicates the transition from acquisition training (sessions 1–8) to extinction training (sessions 9–13, indicated by open symbols). Significant ($P<0.05$) pairwise differences are indicated by symbols: * BSR vs. FS+BSR; \times FS vs. FS+BSR; + FS vs. BSR. (b) Mean reaction time (\pm SEM) during acquisition sessions (4–8).

interaction [$F_{(6,1,45.7)}=6.1$, $P<0.005$], and main effects of SESSION [$F_{(3,45)}=52.56$, $P<0.001$] and GROUP [$F_{(2,15)}=9.9$, $P<0.005$]. The differences in the CR rate in the three groups decreased over the course of learning. In group FS+BSR, early (sessions 1–3) acquisition was accelerated by the combination of both reinforcers. Tukey post-hoc comparisons confirmed significant differences between the groups during the first three sessions but not during later (sessions 4–8) stages of acquisition. Specifically, group FS+BSR reached significantly higher CR rates compared to group FS during the first session ($P<0.001$) and compared to group BSR during sessions 2 and 3 ($P<0.001$). There was no significant difference between groups BSR and FS observed during the first session of training ($P>0.05$). However, group FS performed better than group BSR ($P<0.005$) during the second and third sessions.

Analysis of reaction time (see Experimental Procedures) was restricted to sessions 4–8, after learning had stabilized (Fig. 2b). Reaction time analysis in the three groups showed no effect of GROUP \times SESSION interaction [$F_{(4,21,31.5)}=1.56$, $P>0.2$], but main effects of SESSION [$F_{(2,1,31.5)}=7.4$, $P<0.005$] and GROUP [$F_{(2,15)}=10.6$, $P<0.005$]. Subjects trained with the combination of both reinforcers showed significantly shorter RTs compared to group FS ($P<0.05$), but not compared to group BSR ($P>0.05$). RTs did not differ between groups FS and BSR ($P>0.05$), except for the last ses-

sion ($P<0.05$). RTs in group FS tended to be longer than in groups BSR and FS+BSR, indicating a speed-up of RT with the involvement of rewarding stimuli. The evaluation of mean number of intertrial crossing across the eight acquisition sessions showed no significant differences between the three groups [$F_{(2,105)}=1.51$, $P=0.25$].

The last session of acquisition (session 8) served as a baseline for extinction analysis. All groups showed a clear extinction of the conditioned hurdle crossing over sessions (Fig. 2a). Main effects for factor SESSION [$F_{(2,6,39.1)}=55.5$, $P<0.0001$] and for factor GROUP [$F_{(2,15)}=6.2$, $P<0.05$] but no significant GROUP \times SESSION interaction [$F_{(5,21,39.1)}=1.66$, $P>0.05$] was found. Tukey post hoc analysis revealed that the main effects were carried by differences between group FS+BSR on the one hand and groups FS and BSR on the other while groups FS and BSR did not differ significantly from each other ($P>0.05$).

Experiment 2. Effects of omission of one reinforcer after completed training with a combined appetitive-aversive reinforcer (continuous reinforcement schedule)

Because of the qualitatively different nature of appetitive and aversive reinforcers, quantitative comparisons between their effects are not trivially achieved. For example, a direct comparison of the effect of reinforcer (appetitive, aversive or combined appetitive-aversive) on

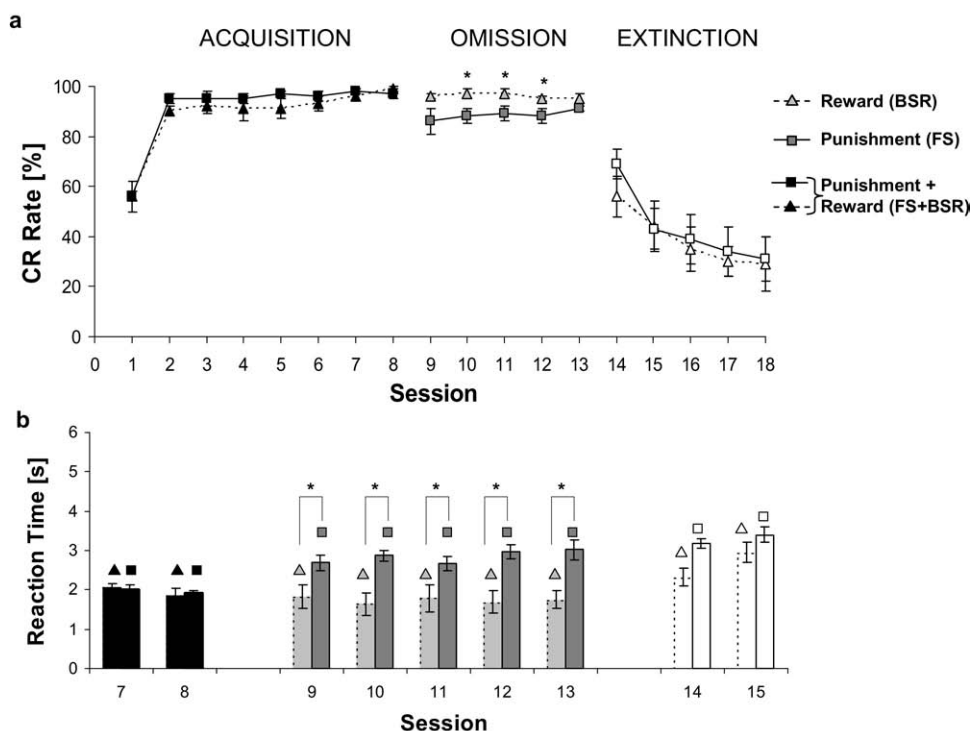


Fig. 3. Effects of removing one reinforcer in animals trained with the combined appetitive-aversive reinforcement using a continuous reinforcement schedule. After training with the combined reinforcers (black filled) during sessions 1–8, BSR was omitted in one group (retaining FS reinforcement, $n=6$, dark grey squares) and FS was omitted in the other (retaining BSR reinforcement, $n=6$, light grey triangles) in sessions 9–13. In sessions 14–18 the previously retained reinforcer was also omitted (extinction). (a) Mean CR rates (\pm SEM) for the two experimental groups (dotted vs. continuous lines) plotted against training session (* indicates significant group differences, $P<0.05$). (b) Mean reaction time (\pm SEM) for acquisition sessions (7–8), omission sessions (9–13) and extinction sessions (14–15).

the acquisition of a CR in Experiment 1 was made possible by the adjustment of reinforcement parameters to values that would produce indistinguishable asymptotic CR rates. This, of course precluded an analysis of potential effects of three reinforcement types on the retention of an already acquired CR. To study potential effects of the different reinforcer type on the retention of an already acquired CR, in a second experiment, we trained two groups of animals to a high level of perfor-

mance using combined appetitive-aversive training (session 1–8), and then omitted either the appetitive or the aversive reinforcer while continuing the other (session 9–13). Subsequently (session 14–18), the previously remaining reinforcer was also omitted, effectively resulting in an extinction session. This omission approach was carried out in two versions, one employing continuous reinforcement (Experiment 2) and one employing partial reinforcement (Experiment 3).

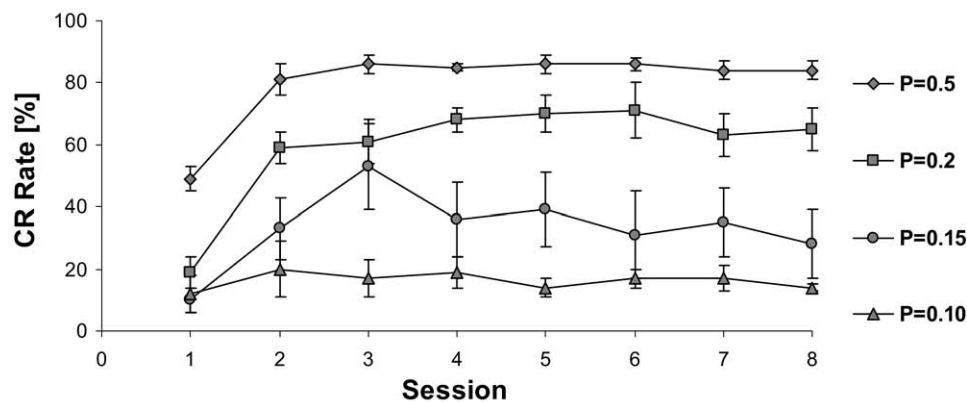


Fig. 4. Mean CR rate \pm SEM using different probabilities of FS presentation following misses. The probabilities studied were $P=0.5$ ($n=5$), $P=0.2$ ($n=5$), $P=0.15$ ($n=4$) and $P=0.10$ ($n=3$). This experiment was a pilot study for Experiment 3. FS probability of 0.15 produced asymptotic CR rates between 30% and 50% and was selected as the probability for the combined appetitive-aversive reinforcement in Experiment 3.

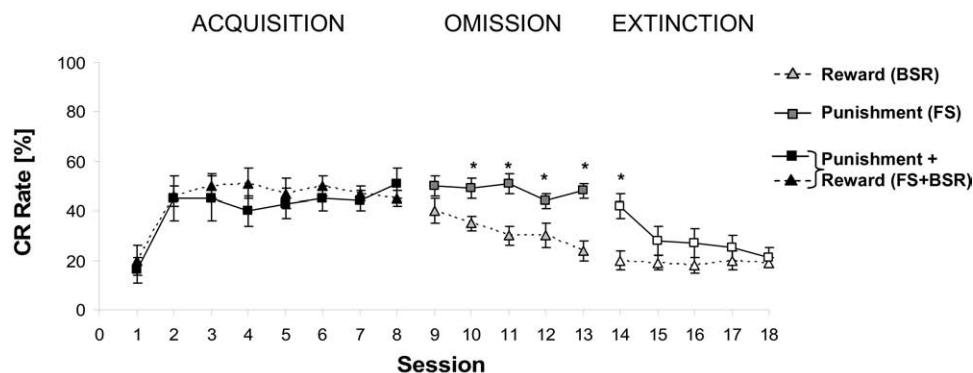


Fig. 5. Effects of removing one reinforcer in animals trained with the combined appetitive-aversive reinforcers under partial reinforcement. Mean CR rates (\pm SEM) for both groups were plotted against training session. After training with both reinforcers (black filled) during sessions 1–8, BSR was omitted in one group (retaining FS, $n=5$, dark grey squares) and FS was omitted in the other (retaining BSR, $n=5$, light grey triangles) in sessions 9–13. In sessions 14–18 the previously retained reinforcer was also omitted (extinction). Asterisks indicate significant group differences, $P<0.05$.

Two experimental groups were trained in an identical fashion using combined appetitive-aversive reinforcers (FS+BSR) during acquisition training (sessions 1–8; Fig. 3a). Consequently, the statistical analysis of the CR rates demonstrated that the two groups acquired the same level of performance at the same rate: the ANOVA revealed a main effect for SESSION [$F_{(2,9,29,2)}=114.3$, $P<0.001$], but not for GROUP [$F_{(1,10)}=1.28$, $P>0.05$] or the interaction GROUP \times SESSION [$F_{(2,9,29,2)}=1.02$, $P>0.05$]. Likewise, analysis of reaction time indicated no significant differences

between the experimental groups (GROUP \times SESSION [$F_{(3,4,34,4)}=0.6$, $P>0.5$]; SESSION [$F_{(3,4,34,4)}=22.7$, $P<0.001$]; and GROUP [$F_{(1,10)}=0.005$, $P>0.5$]).

Following omission of either FS or BSR (sessions 9–13; Fig. 3a), CRs continued at high rates which remained constant across all five sessions, reflected by an absence of a SESSION effect [$F_{(2,6,26,5)}=0.33$, $P>0.5$] and of a GROUP \times SESSION interaction [$F_{(2,6,26,5)}=0.70$, $P>0.5$]. However, a main effect for factor GROUP was found [$F_{(1,10)}=8.9$, $P<0.05$]. The omission of BSR caused

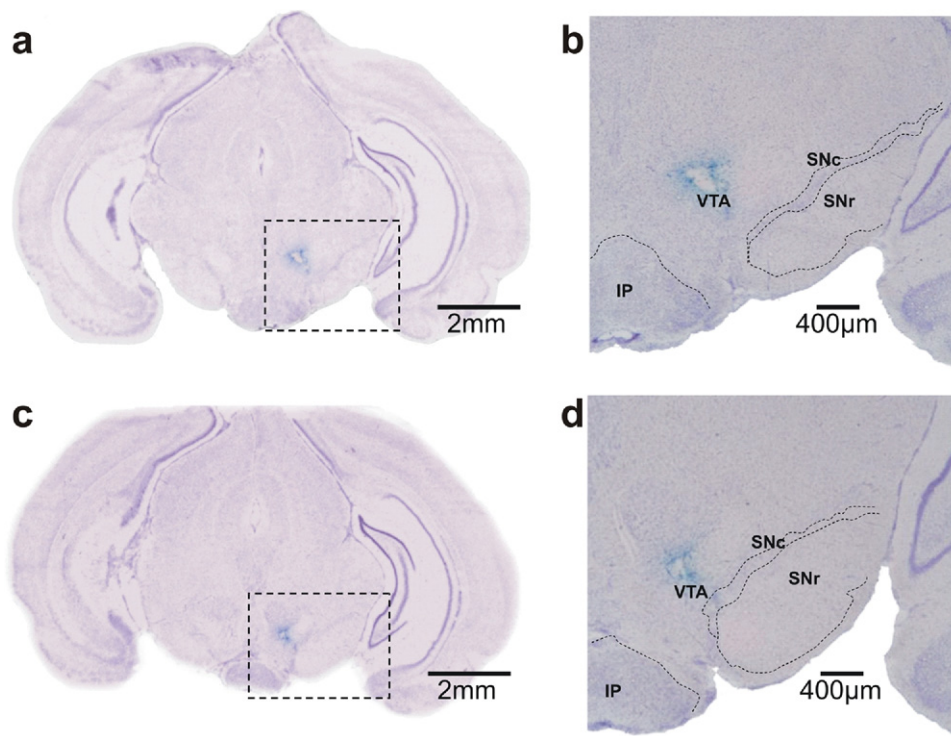


Fig. 6. Localization of stimulation sites for BSR using histological analysis. (a, c) Exemplary photomicrographs of two coronal sections ($40\ \mu\text{m}$) from approximately 2.6 and 2.8 mm posterior to bregma. The sections were stained with Prussian Blue followed by Nissl staining. Note the blue color reaction product around the electrode location in the stimulated hemisphere. (b, d) Magnification of the area surrounding the stimulation site in the ventral tegmental area from panels a and c (rectangular areas). Anatomical structures: IP, interpeduncular nucleus; SNc, substantia nigra pars compacta; SNr, substantia nigra pars reticulata; VTA, ventral tegmental area.

a slight drop in responding whereas the omission of FS had no effect. Paired *t*-tests were used to compare the last session of acquisition and the first session of omission in each group. A non-significant difference in BSR (FS omitted) ($t=1.75$, $df=5$, $P=0.14$) and significant difference in group FS (BSR omitted) ($t=2.61$, $df=5$, $P=0.04$) were found. Significant differences in reaction time (Fig. 3b) were found during the omission sessions (GROUP \times SESSION [$F_{(2,46,24,6)}=2.73$, $P>0.05$]; SESSION [$F_{(2,46,24,6)}=0.90$, $P>0.05$]; and GROUP [$F_{(1,10)}=15.04$, $P<0.005$]). In summary, an increase in reaction time was observed for omission of reward but not for omission of punishment.

The analysis of the extinction sessions (session 14–18; Fig. 3a) revealed clear extinction (main effect of SESSION: [$F_{(2,7,27,7)}=2.07$, $P<0.001$]), but no differences between the two experimental groups (GROUP: [$F_{(1,10)}=0.14$, $P>0.05$]; GROUP \times SESSION [$F_{(2,7,27,7)}=0.93$, $P>0.05$]). These results suggest that, after omission of one reinforcer, the strength of the conditioning was the same during extinction.

Experiment 3. Effects of omission of one reinforcer after completed training with a combined appetitive-aversive reinforcer (partial reinforcement schedule)

The same type of experiments as described above was repeated using partial reinforcement because we speculated that the effect of omitting the internal reward during the continuous reinforcement procedure could be due to the reinforcer history. Specifically, the small decrease in CR rate in group FS (omission of BSR) not seen in group BSR (omission of FS) could simply reflect that, when highly trained animals showing significantly more hits than misses, omission of BSR during hits implied a greater change in reinforcement than omission of FS during misses. We therefore adjusted the probability of FS and BSR association (partially reinforced), such that the final performance of animals trained with the combined appetitive-aversive reinforcers was near 50% CR rate.

Under the partial reinforcement condition using combined appetitive-aversive reinforcers (Fig. 5), significant acquisition (sessions 1–8) of the CR was observed (SESSION: [$F_{(3,0,24,1)}=12.5$, $P<0.001$]) that was expectedly indistinguishable for the two experimental groups (GROUP: [$F_{(1,8)}=0.3$, $P>0.5$]; GROUP \times SESSION: [$F_{(3,24,1)}=0.71$, $P>0.5$]).

Subsequent omission of one reinforcer under partial reinforcement (session 9–13) led to clear changes in the CR rate development (SESSION: [$F_{(2,6,21,0)}=5.8$, $P<0.05$]) and significant group differences (GROUP: [$F_{(1,8)}=11.4$, $P<0.05$]; GROUP \times SESSION: [$F_{(2,6,21,0)}=3.6$, $P<0.05$]). These results demonstrate that under partial reinforcement conditions, omission of FS for misses decreased the CR rate but omission of BSR for hits did not. Analysis of the extinction curves (session 14–18; Fig. 5) revealed a main effect for SESSION [$F_{(2,8,23,1)}=8.7$, $P<0.005$] and GROUP \times SESSION remained significant [$F_{(2,8,23,1)}=6.54$, $P<0.005$] but not for GROUP [$F_{(1,8)}=2.82$, $P>0.1$].

DISCUSSION

In essence, we found that acquisition of a new behavior, tone-conditioned hurdle crossing, can be learned either by appetitive or aversive reinforcers and is facilitated by the interaction of the two as predicted by an equivalence hypothesis. This facilitated learning, characterized by very fast acquisition and close to 100% performance level, is extremely unusual in a 2-way shuttle-box paradigm, and was only described in rats for a particular genetically-selected strain (Aguilar et al., 2004; Corda et al., 2005). Taken together, our results demonstrate a strong learning improvement caused by the functional interaction between appetitive and aversive reinforcers. The acquisition of the CR under the influence of combined appetitive-aversive reinforcers (Fig. 2a; sessions 1–8) suggests that the successful avoidance effect associates with the stimulation effect of intrinsic reward, therefore leading to faster acquisition. Our present findings also support previous studies (Tanimoto et al., 2004; Stark et al., 2004) indicating that the termination of aversive stimuli itself is rewarding. Therefore, addition of intrinsic reward to the relief from punishment in a time-dependent manner accelerated the performance. Although the nature of interaction of appetitive reward and relief from aversive influences is not explicitly known, evidence of some functional convergence of the two influences in the human orbitofrontal cortex has been obtained (Kim et al., 2006).

After reaching an asymptotic level of performance with multiple trials (i.e., at the stage of retrieval of conditioning) a correlate of the strength of the consolidated memory can be obtained from the extinction phenomenon in non-reinforced trials. In general, avoidance responses with aversive reinforcers have appeared more resistant to extinction than appetitive conditioning (Mackintosh, 1983), presumably because of underlying fear resulting from the first classical tone-shock conditioning stage of instrumental learning. In our results (Fig. 2a), however, no significant difference in extinction was observed between the groups trained with appetitive or aversive reinforcer. This is probably due to differences in experimental procedures. In our experiment direct stimulation of the internal reward system was used instead of a natural reward, e.g. food, which nevertheless suggests that internal reward-driven learning in principle, can be as resistant to extinction as avoidance learning.

As already suggested by the higher final level of performance, the strength of conditioning as revealed by extinction was much higher in the group trained with combined reinforcers (Fig. 2a). This group showed high resistance to extinction probably because response feedback by VTA stimulation for avoidance creates powerful long-lasting stimulus-response habits (Graybiel, 2008). Even partial extinction required several sessions, and it is not obvious from the available data whether full extinction can be obtained. The strength of conditioning in this group may be explained by an additive effect in the internal reward system. We have previously found in the FS-reinforced avoidance conditioning that a transitory strong increase of dopamine level occurs in medial prefrontal cortex when

after perfection of the initial escape strategy the animals experience a number of first successful FS avoidances (Stark et al., 2004). This has been interpreted as a correlate of relief from punishment. As the successful avoidance trials are also rewarded with brain stimulation there should be some additive effect in terms of dopamine release. Considering dopamine release as a predictive measure of the speed of learning (Stark et al., 1999) and final memory consolidation in the shuttle-box (Stark et al., 2004) the comparison of the present three experiments might allow even some inferences about recruitment of the dopaminergic system. Although the final level of performance was (by design) indistinguishable between groups FS and BSR, the latter was slowest in developing this CR level and fastest in extinction. This might suggest that the dopamine release during relief of punishment in the FS-reinforced group was larger or at least more effective. This interpretation is supported by the fact that dopamine seems to play different roles for learning processes in different brain areas (cf. Gallistel, 2006; Hernandez et al., 2006; Day et al., 2007). Also, dopamine cannot govern all motivational processes, thus the relative engagement of other systems should be explored. For example, an interaction of the serotonergic raphe nuclei (Daw et al., 2002) and the lateral habenula (Matsumoto and Hikosaka, 2007) with the midbrain dopamine system may encode motivational states during opponent appetitive and aversive processes (Seymour et al., 2005).

The purpose of the selective reinforcer omission experiments in the groups trained with combination of both reinforcers after reaching stable performances was to examine how the two reinforcers interact at high performance levels to maintain that level. While omission of FS did not change the maximum performance level, omission of rewards led to a drop of performance (Fig. 3a). At first glance this does not seem astonishing from the point of view of a number of events. During asymptotic performance, there was only occasional FS while almost all trials were reinforced by BSR. Nevertheless in view of the more effective influence of FS on acquisition and retention during extinction (see above), some drop of performance might be expected. Taken together, the effects show that at high performance, omission of one or the other reinforcer must lead to highly asymmetric results and suggest that reward is more important to maintain high levels of performance than punishment. It should be noted that this interpretation, while addressing the mechanistic level of events, is in accordance with Konorski's psychological model of opponent processes and the motivational states between them: namely, predicted appetitive stimuli (hope), predicted aversive stimuli (fear), absence of predicted appetitive stimuli (frustration), and absence of predicted aversive stimuli (relief) events (for recent review see Seymour et al., 2007). Thus, we provide an experimental framework of results which supports the current motivational hypothesis that argues for the functional relevance of aversive inhibitors (relief from punishment) and appetitive exciters (expectation of reward). The conclusion from this part of the study suggests that punishment accelerated the initial acquisition while reward maintained the vigor of learned re-

sponses. However, the drop in performance followed by the omission of reward for hits was reversible and approached the previous level.

The decreased CR rate, observed after omission of reward in animals trained with the combined appetitive-aversive reinforcer, might be attributable to the reinforcer history during continuous reinforcement, namely, the predictability of reward. Therefore, partial conditioning allowed us to manipulate the effect of reinforcer history by violating the reinforcer expectation in an unpredictable fashion. The comparison of results from omission of one reinforcer in continuous and partial reinforcement highlights the fact that the strength of reinforcers during fully predicted continuous and partial reinforcement conditions are different (Figs. 3a and 5). Under partial reinforcement, omission of FS decreased the performance and omission of BSR had no effect in contrast to continuous reinforcement where the opposite effect was found. The conclusions derived from omitting one reinforcer suggest that in a fully predicted, continuously reinforced environment, reward was found to be more effective, at least at the beginning of the omission phase, but, in an unpredictable, partially reinforced situation, punishment is more effective—in each case due to different frequencies of occurrences of rewards and punishments.

Despite the substantial progress in our understanding of appetitive and aversive reinforcement, the old question of which brain systems integrate them (e.g. Valenstein and Valenstein, 1964) needs further exploration (Leknes and Tracey, 2008). Orbitofrontal cortex successfully encodes the expected aversive and appetitive outcomes during learning (Schoenbaum et al., 1998). Subsequent to a demonstration with event-related fMRI that medial parts of orbitofrontal cortex in humans are activated by an appetitively conditioned CS and lateral parts by an aversively conditioned CS (Gottfried et al., 2002), results of a visual conditioning task showed that appetitively and aversively reinforced activations map onto the same area in medial orbitofrontal cortex, suggesting that they interact there (Kim et al., 2006). From a clinical perspective, it was reported that Parkinson patients in a trial-and-error based learning experiment with reward or no reward feedback, if devoid of L-dopa medication, learned more from mistakes, while patients with high levels of medication learned more from successes (Frank et al., 2004).

CONCLUSION

We found that acquisition of a new behavior, tone conditioned hurdle crossing, can be learned by using either appetitive or aversive reinforcer and is facilitated by the interaction of the two. Addition of brain stimulation reward to relief from punishment accelerated performance. Both reinforcers operated together to drive learning, but their effects were still dissociable. In conclusion, we demonstrate that training by “carrot and stick” combined is a superior method to facilitate learning and to counteract extinction. It is noteworthy that, in the combination of both reinforcers, reward and relief from punishment are only equivalent in the sense of a convergent learning effect but cannot replace

each other for reaching maximal learned performance. Our finding of converging and dissociable effects on acquisition and maintenance of learning has implications for optimizing educational and rehabilitational strategies.

Acknowledgments—We thank Dr. Jason Shumake, The University of Texas at Austin, for critical discussion and help during preparation of the manuscript as well as Lydia Löw and Kathrin Ohl for technical assistance during the experiments. We thank Dr. Eike Budinger, Achim Engelhorn, and Dr. Jürgen Goldschmidt for their help with the histological preparations. The study was supported by grants from the Deutsche Forschungsgemeinschaft (DFG SFB-779), the Land Sachsen Anhalt (XN3621C/1005M) and the European IST Programme (FP6-0027787).

REFERENCES

- Aguilar R, Gil L, Fernández-Teruel A, Tobena A (2004) Genetically-based behavioural traits influence the effects of shuttle box avoidance overtraining and extinction upon intertrial responding: a study with the Roman rat strains. *Behav Processes* 66:63–72.
- Aldavert-Vera L, Costa-Miserachs D, Massanes-Rotger E, Soriano-Mas C, Segura-Torres P, Morgado-Bernal I (1997) Facilitation of a distributed shuttle-box conditioning with posttraining intracranial self-stimulation in old rats. *Neurobiol Learn Mem* 67:254–258.
- Baron A, DeWaad RJ, Lipson J (1977) Increased reinforcement when timeout from avoidance includes access to a safe place. *J Exp Anal Behav* 27:479–494.
- Bull JA (1970) An interaction between appetitive pavlovian CSs and instrumental avoidance responding. *Learn Motiv* 1:18–26.
- Carder B (1970) Lateral hypothalamic stimulation and avoidance in rats. *J Comp Physiol Psychol* 71:325–333.
- Carr KD, Coons EE (1982) Rats self-administer nonrewarding brain stimulation to ameliorate aversion. *Science* 215:1516–1517.
- Castro-Alamancos MA, Borrell J (1992) Facilitation and recovery of shuttle box avoidance behavior after frontal cortex lesions is induced by a contingent electrical stimulation in the ventral tegmental nucleus. *Behav Brain Res* 50:69–76.
- Cordea MG, Piras G, Lecca D, Fernández-Teruel A, Driscoll P, Giorno O (2005) The psychogenetically selected Roman rat lines differ in the susceptibility to develop amphetamine sensitization. *Behav Brain Res* 157:147–156.
- Cox VC, Valenstein ES (1965) Attenuation of aversive properties of peripheral shock by hypothalamic stimulation. *Science* 149:323–325.
- Day JJ, Roitman MF, Wightman RM, Carelli RM (2007) Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. *Nat Neurosci* 10:1020–1028.
- Daw ND, Kakade S, Dayan P (2002) Opponent interactions between serotonin and dopamine. *Neural Netw* 15:603–616.
- Dickinson A, Pearce JM (1977) Inhibitory interactions between appetitive and aversive stimuli. *Psychol Bull* 84:690–711.
- Dickinson A, Dearing MF (1979) Appetitive-aversive interactions and inhibitory processes. In: *Mechanisms of learning and motivation* (Dickinson A, Boakes RA, eds), pp 203–231. Hillsdale, NJ: Erlbaum.
- Dickinson A, Mackintosh NJ (1978) Classical conditioning in animals. *Annu Rev Psychol* 29:587–612.
- Dickinson A (1977) Appetitive and aversive interactions: super conditioning of fear by an appetitive CS. *Q J of Exp Psychology* 29:71–83.
- Dickinson A (1976) Appetitive-aversion interactions: facilitation of aversive conditioning by prior appetitive training in the rat. *Anim Learn Behav* 4:416–420.
- Dinsmoor JA (2001) Stimuli inevitably generated by behavior that avoids electric shock are inherently reinforcing. *J Exp Anal Behav* 75:311–333.
- Estes WK, Skinner BF (1941) Some quantitative properties of anxiety. *J Exp Psychol* 29:390–400.
- Frank MJ, Seeberger LC, O'Reilly RC (2004) By carrot or by stick: cognitive reinforcement learning in Parkinsonism. *Science* 306:1940–1943.
- Gallistel CR (2006) Dopamine and reward: comment on Hernandez et al. (2006). *Behav Neurosci* 120:992–994.
- Gottfried JA, O'Doherty J, Dolan RJ (2002) Appetitive and aversive olfactory learning in humans studied using event-related functional magnetic resonance imaging. *J Neurosci* 22:10829–10837.
- Graybiel AM (2008) Habits, rituals, and the evaluative brain. *Annu Rev Neurosci* 31:359–387.
- Hearst E (1960) Simultaneous generalization gradients for appetitive and aversive behavior. *Science* 132:1769–1770.
- Hernandez G, Hamdani S, Rajabi H, Conover K, Stewart J, Arvanitiogiannis A, Shizgal P (2006) Prolonged rewarding stimulation of the rat medial forebrain bundle: neurochemical and behavioral consequences. *Behav Neurosci* 120:888–904.
- Kelleher RT, Cook L (1959) An analysis of the behavior of rats and monkeys on concurrent fixed-ratio avoidance schedules. *J Exp Anal Behav* 2:203–211.
- Kim H, Shimojo S, O'Doherty JP (2006) Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. *PLoS Biol* 4:1453–1461.
- Konorski J, Szwejkowska G (1956) Reciprocal transformations of heterogeneous conditioned reflexes. *Acta Biol Exp* 16:95–113.
- Konorski J (1967) *The integrative activity of the brain: an interdisciplinary approach*. Chicago: Chicago University Press.
- Leknes S, Tracey I (2008) A common neurobiology for pain and pleasure. *Nat Rev Neurosci* 9:314–320.
- Loskota WJ, Lomax P, Verity MA (1974) *A stereotaxic atlas of the Mongolian gerbil (Meriones unguiculatus) brain*. Michigan: Ann Arbor Science Publishers.
- Mackintosh NJ (1983) *Conditioning and associative learning*. New York: Oxford University Press.
- Magoon MA, Critchfield TS (2008) Concurrent schedules of positive and negative reinforcement: differential-impact and differential-outcomes hypotheses. *J Exp Anal Behav* 90:1–22.
- Margules DL, Stein L (1968) Facilitation of Sidman avoidance behavior by positive brain stimulation. *J Comp Physiol Psychol* 66:182–184.
- Matsumoto M, Hikosaka O (2007) Lateral habenula as a source of negative reward signals in dopamine neurons. *Nature* 447:1111–1115.
- Mogenson GJ, Morrison MJ (1962) Avoidance responses to reward stimulation of the brain. *J Comp Physiol Psychol* 55:691–694.
- Morris RGM (1975) Preconditioning of reinforcement properties to an exteroceptive feedback stimulus. *Learn Motiv* 6:289–298.
- Morrison SE, Salzman CD (2009) The convergence of information about rewarding and aversive stimuli in single neurons. *J Neurosci* 29:11471–11483.
- Ohl FW, Wetzell W, Wagner T, Rech A, Scheich H (1999) Bilateral ablation of auditory cortex in Mongolian gerbil affects discrimination of frequency modulated tones but not of pure tones. *Learn Mem* 6:347–362.
- Ohl FW, Scheich H, Freeman WJ (2001) Change in pattern of ongoing cortical activity with auditory category learning. *Nature* 412:733–736.
- Olds ME, Olds J (1962) Approach-escape interactions in rat brain. *Am J Physiol* 203:803–810.
- Reed P, Mitchell C, Nokes T (1996) Intrinsic reinforcing properties of putatively neutral stimuli in an instrumental two-lever discrimination task. *Anim Learn Behav* 24:38–45.
- Rolls ET (2008) *Memory, attention, decision-making*. Oxford, UK: Oxford University Press.
- Scavio MJ (1974) Classical-classical transfer: effects of prior aversive conditioning upon appetitive conditioning in rabbits. *J Comp Physiol Psychol* 86:107–115.
- Schoenbaum G, Chiba AA, Gallagher M (1998) Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. *Nat Neurosci* 1:155–159.

- Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. *Science* 275:1593–1599.
- Seymour B, O'Doherty JP, Koltzenburg M, Wiech K, Frackowiak R, Friston K, Dolan R (2005) Opponent appetitive-aversive neural processes underlie predictive learning of pain relief. *Nat Neurosci* 8:1234–1240.
- Seymour B, Singer T, Dolan R (2007) Neurobiology of punishment. *Nat Rev Neurosci* 4:300–311.
- Stark H, Bischof A, Scheich H (1999) Increase of extracellular dopamine in prefrontal cortex of gerbils during acquisition of the avoidance strategy in the shuttle-box. *Neurosci Lett* 264:77–80.
- Stark H, Rothe T, Wagner T, Scheich H (2004) Learning a new behavioral strategy in the shuttle-box increases prefrontal dopamine. *Neuroscience* 126:21–29.
- Stein L (1965) Facilitation of avoidance behaviour by positive brain stimulation. *J Comp Physiol Psychol* 60:9–19.
- Tanimoto H, Heisenberg M, Gerber B (2004) Event timing turns punishment to reward. *Nature* 430:983.
- Valenstein ES, Valenstein T (1964) Interaction of positive and negative reinforcing neural systems. *Science* 145:1456–1458.
- Wetzel W, Wagner T, Ohl FW, Scheich H (1998) Categorical discrimination of direction in frequency-modulated tones by Mongolian gerbils. *Behav Brain Res* 91:29–39.
- Wetzel W, Ohl FW, Scheich H (2008) Global versus local processing of frequency-modulated tones in gerbils: an animal model of lateralized auditory cortex functions. *Proc Natl Acad Sci U S A* 105: 6753–6758.

(Accepted 6 January 2010)
(Available online 19 January 2010)