Variation in the effectiveness of reinforcement and nonreinforcement in generating different conditioned behaviors

Victor Navarro, Dominic M. Dwyer, Robert C. Honey*

Cardiff University, UK

Abstract

Rat autoshaping procedures generate two readily measurable conditioned responses: During lever presentations that have previously signaled food, rats approach the food well (called goal-tracking) and interact with the lever itself (called sign-tracking). We investigated how reinforced and nonreinforced trials affect the overall and temporal distributions of these two responses across 10-second lever presentations.

In two experiments, reinforced trials generated more goal-tracking than sign-tracking, and nonreinforced trials resulted in a larger reduction in goal-tracking than sign-tracking. The effect of reinforced trials was evident as an increase in goal-tracking and reduction in sign-tracking across the duration of the lever presentations, and nonreinforced trials resulted in this pattern transiently reversing and then becoming less evident with further training.

These dissociations are consistent with a recent elaboration of the Rescorla-Wagner model, HeiDi (Honey, R. C., Dwyer, D.M., & Iliescu, A.F. (2020a). HeiDi: A model for Pavlovian learning and performance with reciprocal associations. Psychological Review, 127, 829–852.), a model in which responses related to the nature of the unconditioned stimulus (e.g., goal-tracking) have a different origin than those related to the nature of the conditioned stimulus (e.g., sign-tracking).

1. Introduction

Pavlovian conditioning and extinction are fundamental learning processes: conditioned responses (CRs) established by pairing a conditioned stimulus (CS) with an unconditioned stimulus (US) wane when the CS is presented alone (Pavlov, 1927). One key feature of the influential Rescorla and Wagner (1972) model was its elegant approach to how reinforced conditioning trials and nonreinforced extinction trials generate these changes in conditioned responding. The model assumed that reinforced trials result in the CS accruing excitatory association strength (i.e., $V_{CS}$ increasing) to an asymptote determined by the US ($\lambda$), and it supposed that nonreinforced extinction trials result in a reduction in $V_{CS}$ contingent on $\lambda$ being lower on nonreinforced than on reinforced trials (p. 76–77, Rescorla & Wagner, 1972; see also, Stout & Miller, 2007). In this way, the model gave formal expression to their general thesis: “Certain expectations are built up about the events following a stimulus complex; expectations initiated by that complex and its component stimuli are then only modified when consequent events disagree with the composite expectation.” (p. 75; Rescorla & Wagner, 1972). Alternative analyses of the effects of extinction trials include the idea that they result in the formation of an inhibitory CS—US association (e.g., Konorski, 1948; Wagner, 1981) or the development of an excitatory CS—no US association (e.g., Konorski, 1967; Pearce & Hall, 1980). Despite their differences, all of these theoretical analyses and their variants assume that there is a simple monotonic relationship between the $V_{CS}$ and conditioned behavior, whether that strength is derived from the effect of reinforced trials alone or the combined effects of reinforced and nonreinforced trials. For example, Rescorla and Wagner (1972) stated that in “providing some mapping of V values into behavior”, it was “sufficient simply to assume that the mapping of Vs into magnitude or probability of conditioned responding preserves their ordering.” This simplifying assumption has been adopted across behavioral and neurobiological studies of learning, but it is undermined by two related observations.

First, a given conditioning procedure can generate conditioned responses that not only reflect the nature of the US, but also responses that...
reflect the properties of the CS (for a review, see Holland, 1984). While it is simple to appreciate how the strength of a directional association from a representation of the CS to that of a US could generate responses that reflect the nature of the US, it is much less clear why such a directional association would generate responses that reflect the nature of the CS. Second, some rats are more likely to exhibit what they have learnt in terms of one form of conditioned responding than another, whereas in other rats the reverse is the case (Flagel et al., 2009; Patitucci et al., 2016). These qualitative individual differences are difficult to reconcile with the simple mapping function adopted by in the Rescorla and Wagner (1972) model, but also by other models of Pavlovian conditioning (e.g., Pearce & Hall, 1980; Pearce, 1994; Rescorla & Wagner, 1972; Stout & Miller, 2007; Wagner, 1981). We now discuss these two observations in a little more detail because they provided the original impetus for the development of a recent model of Pavlovian learning and performance, HeiDI (Honey et al., 2020ab; Honey & Dwyer, 2022). As we will show, this model makes an intriguing prediction about the impact of reinforcement and nonreinforcement on the spatio-temporal distributions of conditioned responding; but first, we consider the impetus for its development.

The fact that conditioning procedures not only generate responses that reflect the nature of the US but also the CS represents a challenge to many formal and general process theories of Pavlovian conditioning (e.g., Holland, 1977, 1984; Timberlake & Grant, 1975; see also, Asratyan, 1965; Pavlov, 1927). For example, consider a rat autoshaping procedure in which the temporary insertion of a lever into a conditioning chamber (the CS) is followed by food delivery (the US). As a result of lever food pairings, the lever comes to elicit approach to the food well (called goal-tracking, a US-oriented response) and interactions with the lever itself (called sign-tracking, a CS-oriented response; Boakes, 1977; Flagel et al., 2009; Patitucci et al., 2016). Although goal-tracking can be readily explained by appealing to a directional CS US association, sign-tracking cannot.

Asratyan (1965) suggested a way to bridge this explanatory gap. He proposed that forward conditioning trials result in the formation of reciprocal CS US and US CS associations (CS US associations). A CS US association would (directly) generate CRs that reflect the nature of the US, while a US CS association would (indirectly) generate CRs that reflect the nature of the CS: Indirectly because associative activation of the CS would be mediated by activation of the US representation. The idea that reciprocal associations form during Pavlovian conditioning has received support from studies of autoshaping in rats (Navarro et al., 2023), and a recent formal model (HeiDI) describes learning rules for implementing this idea, together with performance rules for determining how the reciprocal associations are translated into different conditioned behaviors (Honey et al., 2020ab; Honey & Dwyer, 2021, 2022). The learning rules are simple rationalizations of the Rescorla and Wagner (1972) rules; and HeiDI assumes that upon presentation of the CS the combined associative strengths (Vs) of the CS US and US CS associations (given by: VCS US + [VCS US X CS CS]), is distributed to affect CS-oriented responding and US-oriented responding. The distribution of these two forms of conditioned responding upon presentation of the CS is held to reflect the relative perceived intensities of the CS (which is present) and the US (which is retrieved); with the intensity of the retrieved US being a function of the strength of the association between the CS and US. Under these conditions, when the perceived intensity of the CS is greater than that of the (retrieved) US, then CS oriented behavior dominates US-oriented behavior; and when the perceived intensity of the US is greater than that of the CS, then the reverse is the case. The latter assumption provides a potential basis for individual differences in the biases of rats to exhibit what they have learned predominantly in terms CS or US-oriented responding (e.g., Flagel et al., 2009; Patitucci et al., 2016). Once it is assumed that the perceived intensities of the CS and US vary across animals, some will be more likely to express what they have learnt as CS-oriented behavior, others as US-oriented behavior, and the remainder will show similar levels of the two behaviors (see Honey et al., 2020c). Qualitative individual differences in the form of conditioned responding are quite beyond the Rescorla and Wagner (1972) model and other extant models (e.g., Pearce, 1994; Pearce & Hall, 1980; Stout & Miller, 2007; Wagner, 1981). However, the analysis provided by HeiDI for the generation of different forms of conditioned responding also makes an intriguing prediction, at the group level, about the impact of nonreinforced extinction trials on the distribution of CS and US-oriented behaviors.

According to HeiDI, nonreinforced extinction trials reduce the efficacy of the CS US association, and this will have two consequences. First, it will reduce the combined effect of the reciprocal associations (i.e., VCS US + [VCS US X US CS] will decrease). Second, it will change the likelihood that the resulting combined value will be evident in CS and US-oriented behavior: Because while the perceived intensity of the CS is not assumed to change as a result of extinction trials, there will be a reduction in the perceived intensity of the (retrieved) US. Thus, upon presentation of the CS, the perceived intensity of the US is given by the net associative strength of the CS, which will decline as a result of extinction trials. Extinction trials should therefore increase the likelihood that the net effects of learning will be evident as CS rather than US-oriented behavior. That is, unlike the Rescorla and Wagner model (1972; see also, e.g., Pearce, 1994; Pearce & Hall, 1980; Stout & Miller, 2007; Wagner, 1981) HeiDI not only provides a basis for reinforcement to generate different conditioned behaviors, but also predicts that reinforcement and nonreinforcement will have different effects on these behaviors.

While it is well established that autoshaping trials in rats generate US-oriented responding (i.e., goal-tracking) and CS-oriented responding (i.e., sign-tracking; e.g., Flagel et al., 2009; Patitucci et al., 2016), there is limited evidence concerning the impact of nonreinforced trials on these two forms of responding. There is some evidence that the effects of nonreinforced trials differ in rats who are predisposed to engage in goal-tracking rather than sign-tracking (Ahrens et al., 2016; Fitzpatrick et al., 2019; see also, Beckmann & Chow, 2015). For example, Ahrens et al. (2016) showed that extinction trials result in a greater reduction in goal-tracking, in rats prone to goal-track, than it does sign-tracking, in those prone to sign-track (see also, María-Íñigo et al., 2023). Additional evidence suggests that changing the reinforcement contingencies can have a more dramatic effect on goal-tracking than sign-tracking. For example, Iliescu et al. (2018; Experiment 1) showed that after a discrimination in which one lever was reinforced and another nonreinforced, a reversal of

---

2 Wagner’s (1981, p. 20) description of the SOP model focused on the formation of directional (excitatory and inhibitory) linkages from the CS node to the US node, but also noted the following: ‘‘However, it should be understood that what will be said about the development of such linkages is perfectly general so as to imply additional directional linkages from what is here called the US node to that called the CS node.’’ While the model was applied to the effects of backward conditioning (i.e., US CS pairings; p. 32) on directional linkages from the CS node to the US node, there was no analysis of the implied directional linkages from the US node to the CS node.

3 The simpler proposal that CS-oriented conditioned behavior is related to the product of the reciprocal associations (i.e., CS US US CS) while US-oriented conditioned behavior is related to the strength of the CS US association (cf. Asratyan, 1965) provides no basis for CS-oriented behavior to exceed US-oriented behavior (see also, Honey et al., 2020a, p. 834).
the contingencies resulted in a more rapid decrease in goal-tracking than sign-tracking to the reinforced lever. This difference was evident in rats prone to engage in either goal-tracking or sign-tracking. Taken together, these results suggest that reinforced and nonreinforced trials might have dissociable effects on goal-tracking and sign-tracking; a dissociation that would be consistent with the prediction derived from HeiDI (Honey et al., 2020a).

The two experiments reported here examined the impact of reinforced and nonreinforced trials on the overall distribution of goal-tracking and sign-tracking, and the distribution of these two responses across successive epochs of the CS. Previous research has shown that during reinforced trials there is a redistribution of these two responses across the CS, with goal-tracking increasing and sign-tracking declining (Iliesscu et al., 2020). Iliesscu et al. showed that this finding is consistent with HeiDI’s performance rule given the (additional) assumption that the perceived intensity of the CS declines across its duration while that of the US does not. The concurrent examination of the overall levels of the two CRs during reinforced and nonreinforced trials, and their distribution across the CS represents the requisite level of granularity to both characterize the impact of these trials, and to assess predictions derived from HeiDI. The basis for these predictions together with their fits to the data will be presented formally once Experiments 1 and 2 have been presented.

2. Experiment 1

Rats first received conditioning sessions in which presentations of one lever were followed by the delivery of a food pellet and presentations of another lever that were not followed by food.\(^4\) The rats then received extinction sessions in which the presentations of both levers were not followed by food. The conditioning procedures generate marked levels of both goal-tracking and sign-tracking (see Iliesscu et al., 2018; Patitucci et al., 2016). Our primary interest here was in how the overall and temporal distributions of goal-tracking and sign-tracking are affected by reinforced and nonreinforced trials.

2.1. Methodology

2.1.1. Subjects and apparatus

Thirty-two naïve male Lister Hooded rats (mean ad lib weight = 334 g; range: 302–365 g; supplied by Envigo, UK) were used. They were housed in groups ranging from two to four in standard cages and maintained on a 12-hr/12-hr light/dark cycle (lights on at 7 a.m.) and were maintained between 85 % and 95 % of their ad-lib weights by giving them restricted access to food at the end of each day in their home cages. Rats had continuous access to water in these cages. The research was conducted in accordance with the Home Office regulations, under the Animal (Scientific Procedures) Act 1986 and the authority of PPL number PP3468526 granted to D. M. Dwyer.

Sixteen identical conditioning boxes (30 × 24 × 21 cm: H × W × D; Med Associates, Georgia, VT) were used. Each box was placed in a sound-attenuating shell that incorporated a ventilation fan that maintained the background noise at 68 dB(A). The boxes had two aluminum side walls, with front walls, back walls, and ceilings made from clear acrylic. The floor of each box was formed from 19 steel rods (4.8 mm diameter, 16 mm apart) placed above a stainless-steel tray. Food pellets (45 mg; LabDiet, St. Louis, MO, USA) were delivered to a food well (aperture: 5.3 × 5.3 cm), which was recessed at floor level in the center of the left wall. The food well was equipped with infrared detectors, which upon interruption (e.g., by a rat’s snout in the food well)

\(^4\) The results from the conditioning stage were reported by Navarro et al. (2023) as group Forward, and these results were contrasted with a group that received interspersed forward and backward conditioning trials. The results from the extinction stage were not previously reported.
2.1.4. Results

Fig. 1 depicts the levels of goal-tracking and sign-tracking averaged across the last two conditioning sessions, for the reinforced and non-reinforced levers. Rats successfully acquired the discrimination, with both goal-tracking and sign-tracking rates for the reinforced lever being significantly higher than those for the nonreinforced lever (MPD = 0.39, 95 % HDI = [0.30, 0.48], 0 % in ROPE, and MPD = 0.21, 95 % HDI [0.13, 0.30], 0 % in ROPE, respectively). Additionally, goal-tracking was stronger than sign-tracking for both reinforced and non-reinforced levers (MPD = 0.17, 95 % HDI = [0.05, 0.30], 0 % in ROPE and MPD = 0.04, 95 % HDI = [0.02, 0.06], 0 % in ROPE, respectively).

Fig. 2 shows the levels of goal-tracking and sign-tracking during the final session of conditioning (C16; instead of the final 2-session block shown in Fig. 1) and the levels of these two responses to the previously reinforced lever across 6 daily extinction sessions (E1-E6); the levels of both responses to the previously nonreinforced lever remained very low throughout extinction sessions, and will not be subject to further analysis. During the first extinction session, goal-tracking levels to the previously reinforced lever significantly decreased relative to the last conditioning session (MPD = 0.14, 95 % HDI = [0.06, 0.24], 0 % in ROPE), whereas sign-tracking levels significantly increased relative to the same session (MPD = 0.15, 95 % HDI = [0.05, 0.26], 0 % in ROPE). As a result, sign-tracking levels were significantly higher than the goal-tracking levels during the first extinction session (MPD = 0.19, [0.07, 0.31]), a pattern that was opposite to that seen by the end of training (see Fig. 1). In the remaining 5 extinction sessions, both types of response decreased relative to the last conditioning session, with the levels of sign-tracking during the second extinction session not differing from those seen during the last conditioning session (MPD = 0.01, 95 % HDI = [-0.08, -0.09], 34.03 % in ROPE). Despite this progressive decrease in responding, sign-tracking remained significantly higher than goal-tracking in every extinction session (smallest MPD = 0.05, on session 5) except the last (MPD = 0.04, 95 % HDI = [0.01, 0.06], 1.46 % in ROPE).

Fig. 3 shows the distribution of goal-tracking and sign-tracking across 2-s epochs of the 10-second lever presentations, during the final conditioning session and first extinction session. During the last conditioning session (upper panel), the first two epochs of the lever led to more sign-tracking than goal-tracking (smallest MPD = 0.17, 95 % HDI = [0.09, 0.26], 0 % in ROPE; first CS epoch), but in the remaining three epochs, sign-tracking steadily decreased and goal-tracking steadily increased, resulting in stronger goal-tracking than sign-tracking (smallest MPD = 0.25, 95 % HDI = [0.12, 0.39], 0 % in ROPE; third CS epoch). These distinct patterns of results replicates that reported by Iliescu et al. (2020). During the first extinction session (lower panel), the patterns of sign-tracking and goal-tracking were less marked: Sign-tracking was significantly stronger than goal-tracking during the first three CS epochs (smallest MPD = 0.19, 95 % HDI = [0.07, 0.32], 0 % in ROPE; third CS epoch), but both types of response were not significantly different during the final two CS epochs (largest MPD = 0.10, 95 % HDI = [-0.01, 0.22], 5.54 % in ROPE; fifth CS epoch). Relative to the last conditioning session, goal-tracking during the first extinction session was significantly lower on all but the first CS epoch (smallest MPD = 0.13, 95 % HDI = [0.08, 0.19], 0 % in ROPE; second CS epoch), yet sign-tracking was significantly higher on the last three CS epochs (smallest MPD = 0.11, 95 % HDI = [0.05, 0.18], 0 % in ROPE; fifth CS epoch). As a result, the overall levels of goal-tracking were higher during the final conditioning session than in the last extinction session (MPD = 0.12, 95 % HDI = [0.04, 0.21], 0 % in ROPE), but the opposite was true for the overall levels of sign-tracking (MPD = 0.15, 95 % HDI = [0.06, 0.24], 0 % in ROPE).
The results of Experiment 1 show that nonreinforced (extinction) trials have a greater impact on goal-tracking than sign-tracking, and they also alter the distribution of the two responses across the duration of the lever. During conditioning sessions, the goal-tracking CR increased across the CS while the sign-tracking CR decreased, and these differences were less apparent during extinction sessions. Experiment 2 examined the effect of intermixing reinforced and nonreinforced trials on the development of goal-tracking and sign-tracking, and the distribution of these responses across the duration of lever presentations. For rats in group Continuous, all lever presentations were reinforced, whereas for those in group Partial a random half of the lever presentations were reinforced and the remainder were nonreinforced. This design equates the number and distribution of lever presentations across the two groups but means that group Continuous receives double the number of reinforcements as group Partial. The alternative design of equating the number and distribution of reinforcements across groups would mean that group Partial would receive double the number of lever presentations as group Continuous. In both cases, we predict that the effects of intermixing reinforced and nonreinforced trials would result in patterns of conditioned responding similar to those observed during the nonreinforced extinction trials in Experiment 1: Partial reinforcement having a greater impact on goal-tracking than sign tracking, while reducing the difference between goal-tracking and sign-tracking across successive CS epochs. There is evidence that is consistent with the prediction that partial reinforcement has different effects on goal-tracking and sign-tracking. For example, Davey and Cledland (1982) reported a small-scale autoshaping study in which rats received lever presentations that were either always followed by food (for group Continuous; \( n = 5 \)) or were followed by food on half of the presentations and no food on the remainder (for group Partial; \( n = 5 \)). The level of goal-tracking (as measured by food-well entries) was higher in group Continuous than in group Partial, and the reverse was the case for sign-tracking (as measured by contacts with the lever). However, there was no assessment of the distribution of the two types of responses across the duration of the lever (see also, Anselme et al., 2013; Boakes, 1977; Davey et al., 1981; Fuentes-Verdugo et al., 2020; Robinson et al., 2014).

3. Methodology

3.1. Subjects, apparatus and procedure

Thirty-two naïve male Lister Hooded rats (mean ad lib weight = 271 g; range: 229–295 g; supplied by Envigo, UK) were housed and maintained in the same way as in Experiment 1. The apparatus was that used in Experiment 1. Rats first received 2 sessions of training in which they were trained to retrieve food pellets from the food well, with one type of food pellet presented in one session and a second type of food pellet presented in the second session (grain-based or sucrose-based pellets, LabDiet’s 45% STUM and AIN-76A respectively, counterbalanced). Rats then received alternating sessions of training in which the left lever was presented in one session and paired with one type of food pellet and the right lever was presented in the other and paired with the remaining type of food pellet (with lever identity and food type fully counterbalanced). This arrangement meant that we could assess the efficacy of the two pellet types in generating goal-tracking and sign-tracking. For rats in group Continuous (\( n = 16 \)) the 20 presentations of a given lever within a session were followed by food, whereas for those in group Partial (\( n = 16 \)), half the lever presentations were followed by food and the remainder were not; with no more than 2 successive reinforced or non-reinforced trials. Because the food pellet type did not interact with the effects of principal interest here (i.e., group, epoch, and response), we simply pooled over this factor in the analysis that follows. Other details of Experiment 2 were the same as Experiment 1.

3.1.2. Results

Preliminary analyses showed that the identity of the food (grain-based or sucrose-based pellets) only had a main effect on the results, with grain-based pellets leading to higher goal- and sign-tracking. Therefore, all further analyses were collapsed across levers. Fig. 4 shows the levels of goal-tracking and sign-tracking in groups Continuous and Partial pooled across the left and right levers, which were both either reinforced (group Continuous) or partially reinforced (group Partial). Inspection of the upper panel reveals that the levels of goal-tracking in group Continuous were consistently higher than those in group Partial in all conditioning blocks (smallest MPD = 0.08, 95 % HDI = [0.02, 0.14], 0 % in ROPE; block 2). The lower panels show that the levels of sign-tracking did not significantly differ between the groups in any of the blocks (largest MPD = 0.06, 95 % HDI = [0.01, 0.11], 3.49 % in ROPE; block 5).

It is worth noting that the differences between groups Continuous and Partial, which were evident across blocks of training sessions, did not appear to reflect a difference in the number of reinforcers that they received: Had this been so, then the behaviour of group Partial on later training blocks should have resembled that observed in group Continuous given half the number of blocks. This was clearly not the case.

Fig. 5 shows the distribution of goal-tracking and sign-tracking across lever epochs for the final four, 2-session blocks. Inspection of the upper panel shows that the levels of goal-tracking were higher in group Continuous than in group Partial from the second lever epoch onwards, and these differences were evident in all four blocks (smallest MPD = 0.15, 95 % HDI = [0.07, 0.23], 0 % in ROPE; second lever epoch in block 7). Sign-tracking during the first lever epoch tended to be higher for group Continuous than group Partial (and was significantly so during
block 8). However, sign-tracking was significantly lower in group Continuous than in group Partial during the last two lever epochs, and these differences were evident during the last three blocks of conditioning (smallest MPD = 0.09, 95% HDI = [0.02, 0.15], 0% in ROPE; fourth lever epoch in block 8).

4. General discussion

We examined how reinforcement and nonreinforcement affected the expression of two conditioned responses: goal-tracking and sign-tracking. Conditioning trials generated more goal-tracking than sign-tracking and extinction trials produced a greater reduction in goal-tracking than sign-tracking. This pattern of results was evident whether conditioning and extinction trials occurred in different stages (Experiment 1) or they were intermixed in a partial reinforcement procedure (Experiment 2). These overall differences in the two responses were accompanied by marked variation in their distribution across the duration of the CS: On conditioning trials, the profiles of goal-tracking and sign-tracking differed across the CS (replicating Iliescu et al., 2020), and these profiles were modulated as a result of extinction trials. While the Rescorla and Wagner (1972) model provides no basis for these dissociations, neither does its many successors (e.g., Pearce & Hall, 1980; Pearce, 1994; Stout & Miller, 2007; Wagner, 1981). It might seem tempting to suggest that the overall differences in the two responses reflect goal-tracking simply being more sensitive to reinforcement contingencies – both the presentation and omission of the US – than sign-tracking. However, that suggestion does not explain why extinction alters the distributions of the two responses across the duration of a CS. In the Introduction we briefly outlined an elaboration of the Rescorla and Wagner (1972) model, called HeiDI, with the potential to address these issues (see Honey et al., 2020ab; Honey & Dwyer, 2022). It is now appropriate to fully consider this model and assess its fit to the overall levels and distributions of goal-tracking and sign-tracking in
Experiments 1 and 2.

The HeiDI learning and performance equations are re-presented here using a generalized form of notation and including factorization; with any changes or additions to the original equations being explicitly identified. Like Asratyan (1965), HeiDI assumes that reciprocal associations develop during Pavlovian conditioning trials (CS→US and US→CS), with extinction trials reducing the efficacy of the CS→US association. Trial-based changes in the strengths of the CS→US and US→CS associations are both given by:

$$\Delta v_{i \rightarrow j} = \alpha_i(c_{ij} - \sum_k v_{k \rightarrow j})$$

(1)

where $\Delta v_{i \rightarrow j}$ denotes the change in the association from stimulus $i$ to stimulus $j$, and $\alpha_i$ and $c_{ij}$ denote the perceived intensities of the two stimuli, $K$ is the set containing all stimuli presented on the trial, and $c$ is a constant of 1 in units of associative strength (V), which balances the equations in terms of units of measurement. The $c$ constant is ignored in all remaining equations for the sake of simplicity.

HeiDI assumes that the reciprocal associations between stimuli are combined and distributed into responses that reflect the perceived intensities of the stimuli that are present or associatively activated. The combined associative strength with which stimuli $K$ activate stimulus $j$, $O_{K,j}$, is given by:

$$O_{K,j} = \sum_k v_{k \rightarrow j} \times (1 + \sum_i v_{i \rightarrow k})$$

(2)

Eq. (2) reveals that $O_{K,j}$ is equal to the net forward associations to $j$ plus the net backward associations from $j$ to all stimuli $K$, conditioned by the net forward associations to $j$ itself. The degree to which $O_{K,j}$ is expressed into responding reflecting stimulus $i$’s nature is related to $R_{i,j}$, which is given by:

$$R_{i,j} = O_{K,j} \times \frac{\theta_i}{\sum_i \theta_i}$$

(3)

where $\theta_i$ is a function that depends on whether stimulus $i$ is presented on the trial and $N$ is the set containing all the experimental stimuli. If $i$ is presented ($i \in K$), then $\theta_i$ is equal to the perceived intensity or salience of the stimulus, $\alpha_i$. If $i$ is not presented ($i \notin K$), then $\theta_i$ is equal to the sum of absolute forward associations that the presented stimuli have with stimulus $i, \sum_k |v_{k \rightarrow i}|$. For the current implementation of the model, only stimuli with biological relevance (i.e., those considered to be the US in traditional experimental paradigms) were assumed to support conditioned responses.

According to HeiDI, the perceived intensity of a CS (i.e., $\alpha_{CS}$) does not change as a function of conditioning or extinction trials. However, in the simplifying case in which the set $K$ only contains a single CS and $j$ denotes the US, the strength with which the US is retrieved [its associatively retrieved intensity; Eq. (3)] is increased by conditioning trials, as $v_{CS-US}$ increases, and reduced by extinction trials, as $v_{CS-US}$ tends to zero. This analysis predicts that goal-tracking will be more susceptible to the effects of extinction than sign-tracking: While the reduction in the combined associative strength, $O_{CS-US}$, will be reflected in a decrease in both responses, the reduction in the associatively retrieved intensity of the (absent) US will mean that this combined strength will be less likely to be evident in goal-tracking than sign-tracking. To address how the distribution of the two responses changes across CS duration, we have suggested that $\alpha_{CS}$ declines across the duration of a CS (e.g., due to a short-term adaptation or habituation; see Honey & Dwyer, 2021, 2022, Iliescu et al., 2020; Pavlov, 1927, p. 104; see also, Staddon, 2005; Staddon & Higa, 1999). This suggestion means that $R_{CS-US}$ (and its impact on sign-tracking) should be more evident at the start of a CS than its end, and $R_{US,US}$ (and its impact on goal-tracking) should be more evident at the end of the CS than its start (see Cinotti et al., 2019; Derman et al., 2018; Holland, 1977; Iliescu et al., 2020; Nasser et al., 2015; but see Lee et al., 2018). This analysis predicts that early extinction trials will preferentially reduce the overall levels of goal-tracking relative to sign-tracking, and provides a basis for sign-tracking to decline across epochs and for goal-tracking to increase. It is worth noting that the increase in the goal-tracking CR across a CS represents an example of a
well-known phenomenon called inhibition of delay (Pavlov, 1927) for which there are many potential explanations (see Mackintosh, 1974, pp. 61-62). However, these explanations provide no coherent account for why the sign-tracking CR shows a quite different change across the CS.

The idea that the perceived intensity of a stimulus declines across its duration is captured in the simulations that follow by assuming that the perceived intensity of the CS at each epoch is given by: $\Delta v^i_{t,j} = \alpha^i_j \times (1 - \rho^i_j t)$, for $e > 1$ and where the initial intensity of the CS, $\alpha^i_0$, is the maximum intensity the CS can take (bound between 0 and 1). Within this parametrisation, $\lambda$ determines the rate of decay, and $\rho$ determines the power of the decay function (although both are closely related in the present application). Critically, we further assumed that on a standard conditioning trial, the perceived intensity of the final epoch of the CS forms an association with the US, and there is intensity-based generalization to other epochs of the CS. Evidence supporting this assumption is presented in Honey and Dwyer (2022). To accommodate this assumption, the learning mechanism presented in Eq. (1) becomes conditional on the stimuli presented in epoch $e$, as:

$$\Delta v^i_{t,j} = \alpha^i_j \times (1 - \sum_k \rho^i_k t)$$

Upon this conditionality on epoch $e$, we modify Eq. (2) to enable earlier conditioning trial, the perceived intensity of the final epoch of the CS becomes conditional, CS epoch via intensity-based generalization, as:

$$\alpha^i_j = \frac{\sum_{t=1}^{T} S(a^i_t, \omega^i_{t,j}) \times (1 + \sum_{k} \rho^i_k t)}{\sum_t S(a^i_t, \omega^i_{t,j})}$$

where $T$ is the set containing all stimulus epochs (exclusive), and $S$ is a similarity function comparing the intensity of a given stimulus at two points in time:

$$S(a^i_t, t) = \frac{e^t}{\epsilon^t + |e - t|} \times \frac{t + e - 1}{t + |e - t|}$$

where $\psi$ is a free parameter between 0 and $+\infty$ modulating the operation of the similarity function across the different epochs. With $\psi = 0$, $S(e,t) = 1$, which means that there is perfect generalization between stimulus epochs. As $\psi$ approaches infinity, $S(e,t)$ approaches 0, which means that there is perfect discrimination between stimulus epochs.

Finally, Eq. (3) becomes:

$$R^i_{t,j} = \sum_k \omega^i_k \times \frac{\rho^i_k}{\rho^i_j}$$

where the function $\omega^i_k$ for an absent stimulus is now the similarity-weighted absolute forward associations that point to it, $\sum_k \sum_{t=1}^{T} S(a^i_t, \omega^i_{t,j}, \omega^i_k)$.  

Fig. 6 depicts the fits of the HeiDI model to the data from Experiment 1, and the fits for the same model equipped with response competition (HeiDI + RC). In the HeiDI + RC model, the $\omega^i_{US,US}$ and $\omega^i_{US,CS}$ (Eq. (7); or $R^i_{US}$ and $R^i_{CS}$ for short) quantities inhibit one another in proportion to their relative strength by a factor of $\omega$ (a free parameter), so, $R^i_{CS} = R^i_{CS} - \omega \times R^i_{US}$ and $R^i_{US} = R^i_{US} - \omega \times R^i_{CS}$. We used the calmr package in R (Navarro, 2024) and custom code (see OSF repository) to estimate the parameters that best fitted the mean response rates (across lever epochs and sessions) via maximum likelihood estimation (see supplemental information for additional details). For both models, we estimated the maximal salience value the levers could take on their first 2 s epoch, $\omega^i_C$, and the intensity of the food, $\omega^i_l$. We also estimated the two parameters for the $\omega^i_C$ decay function (rate and shape). Additionally, for HeiDI + RC we estimated $\omega$, the degree to which $R^i_{CS}$ (evident in goal-tracking) and $R^i_{US}$ (evident in sign-tracking) competed. With $\omega > 0$, this competition mechanism amplifies already-existing differences reflected in $R^i_{CS}$ and $R^i_{US}$.

Panel A of Fig. 6 shows the fit of each model of the data on the last conditioning session and the subsequent 6 extinction sessions. Table 1 contains the best-fitting parameters. Both the base and response competition model achieve a close fit of the data (Akaike’s Information Criterion, 3351 and 3355, respectively), capturing the faster extinction of goal-tracking (Fig. 6A) and the temporal dynamics of goal- and sign-tracking (Fig. 6B). However, both models somewhat underestimate the levels of sign-tracking observed at the outset of extinction. The discrimination sensitivity parameter in both models was very close to zero, which meant that the early and later epochs of the reinforced lever had similar associative properties (i.e., both $\theta^i_C$ and $\theta^i_{US}$ were similar across all epochs). This feature allows the early epochs to generate more goal-tracking than sign-tracking, and the later epochs to generate more goal-tracking than sign-tracking (see also, Iliescu et al., 2020); which would not be the case if the sensitivity parameter was, for example, close to 1. Under these conditions, the later epochs of the CS would have greater associative properties than the earlier epochs.

Fig. 7 depicts the corresponding simulations for Experiment 2, with Table 1 showing the best-fitting parameters (which were optimized simultaneously over both groups). Inspection of panel A of Fig. 7 shows that both versions of HeiDI reproduce the same observation: Relative to a CS that is continuously reinforced, a CS that is partially reinforced supports less goal-tracking, but similar levels of sign-tracking (Akaike’s Information Criterion, 9560 and 9564, respectively), capturing the faster extinction of goal-tracking (Fig. 6A) and the temporal dynamics of goal- and sign-tracking (Fig. 6B). Additionally, panel B shows that both models exhibit the increase in goal-tracking and low levels of sign-tracking across the duration of the continuously reinforced CS. Notably, for Experiment 2 the sensitivity parameter was close to 1 for both models; meaning that it was important for the associative properties of the early epochs of the levers to be lower than the later epochs: In this case, the overall divergence between the two responses across the epochs, requires the model to generate increasingly stronger

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Range</th>
<th>HeiDI Exp. 1</th>
<th>HeiDI Exp. 2</th>
<th>HeiDI + RC Exp. 1</th>
<th>HeiDI + RC Exp. 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha^i_l$</td>
<td>US intensity</td>
<td>(0, 1)</td>
<td>0.60</td>
<td>0.52</td>
<td>0.67</td>
<td>0.55</td>
</tr>
<tr>
<td>$\alpha^i_C$</td>
<td>CS intensity on first epoch</td>
<td>(0, 1)</td>
<td>0.61</td>
<td>0.24</td>
<td>0.58</td>
<td>0.25</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>CS intensity decay rate</td>
<td>(0, 1)</td>
<td>0.28</td>
<td>0.06</td>
<td>0.49</td>
<td>0.19</td>
</tr>
<tr>
<td>$\rho$</td>
<td>CS intensity decay power</td>
<td>(0, +\infty)</td>
<td>0.85</td>
<td>0.06</td>
<td>0.38</td>
<td>0.14</td>
</tr>
<tr>
<td>$\Psi$</td>
<td>Discrimination sensitivity</td>
<td>(0, +\infty)</td>
<td>0.06</td>
<td>1.15</td>
<td>0.25</td>
<td>0.98</td>
</tr>
<tr>
<td>$\omega$</td>
<td>Response competition</td>
<td>(0, +\infty)</td>
<td>0.08</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5 The fits presented here excluded the goal-tracking and sign-tracking responses during the first 2 s of the lever (i.e., lever epoch 1), because the lever is being extended into the chamber during this epoch, introducing noise. However, fits obtained using all the epochs (available in the OSF site of the project) do not alter the observations presented here.

6 The models were fitted on CS epoch-level data, but Figs. 6A and 7A present session-level data. Those data were obtained by averaging over CS epoch-level mode responses, not by fitting the models directly to session-level data.
activation of the US representation (i.e., increasing both $\theta_{US}$ and $O_{RC,US}$) across the epochs.

The need to change the sensitivity parameter in order to fit the results from Experiments 1 and 2 might at first seem perplexing. However, insight into this requirement can be gained from considering the procedural differences between the experiments and conceiving the levers to have unique elements, L1 and L2, and common elements, X. In Experiment 1, rats received a true discrimination (L1 and L2 had different outcomes: $L1X$→food and $L2X$→no food), whereas in Experiment 2 rats either received a pseudo-discrimination, in which L1 and L2 were equally likely to be followed by food and no food (for group Partial: $L1X$→food/no food and $L2X$→no food/no food), or L1 and L2 were consistently followed by food (for group Continuous: $L1X$→food and $L2X$→food). These differences will mean that in Experiment 1 the common elements, X, will have less associative strength than in Experiment 2 (cf. Wagner et al., 1968; see also, Honey et al., 2020a). If some of these common elements include the perceived intensity of L1 and L2, then there are grounds for supposing that the contribution of the associative properties of these elements to performance will be less in Experiment 1 than in Experiment 2. To model this speculative analysis would require additional assumptions to be made about how the identities of L1 and L2 are related to their perceived intensity and how these features (i.e., identity and perceived intensity) might be integrated (cf. Honey et al., 2020ab), which is beyond the scope of the current paper.

It is important to note that the parameters we obtained (Table 1) were relatively consistent across experiments for each model, and sometimes across the models themselves, attesting to the identifiability of the models. For example, both models estimated $\alpha_{US}$ to be higher than $\alpha_{CS}$, although we imposed no such restrictions during parameter estimation. That relation between $\alpha_{US}$ and $\alpha_{CS}$ is critical for the model to estimate the distribution of US- and CS-oriented responding through $R_{US}$ and $R_{CS}$. The other, non-associative parameters such as the decay function’s rate and shape, and the magnitude of response competition were also relatively stable. However, while the simulated results presented in Figs. 6 and 7 are qualitatively similar to the results from Experiments 1 and 2, the fits exhibit systematic variation from some aspects of the results. For example, and regardless of the experiment, both HeiDI and HeiDI + RC models have difficulty estimating response levels early during a trial (i.e., overestimating goal-tracking levels and underestimating sign-tracking levels). We believe equipping the models with extra, non-associative parameters (such as baseline rates for goal-tracking and sign-tracking) could resolve these issues but leave such investigations for the future. It is also worth noting that both models assume no inherent differences in either the properties of the responses or in our sensitivity to measuring them. However, goal-tracking and sign-tracking have quite different motoric and physical characteristics, and we cannot know whether our experimental apparatus (e.g., for monitoring food-well activity and lever interactions) is equally sensitive to measuring one response as the other. This issue could be addressed by using a visual stimulus in one food well to signal the upcoming availability of food in a second food well (see Honey et al., 2020c).

To conclude: The observations that the nature and temporal distribution of different conditioned responses vary as a function of conditioning and extinction are not easily reconciled with many formal models of Pavlovian conditioning (e.g., Pearce & Hall, 1980; Pearce, 1994; Rescorla & Wagner, 1972; Stout & Miller, 2007; Wagner, 1981). Admittedly, such models were developed without these complexities in mind. However, our observations are consistent with predictions derived from a formal model, HeiDI, which builds on the fundamental contributions of the Rescorla and Wagner (1972) model. To be more specific, HeiDI adapts the pooled error term introduced by Rescorla and Wagner (1972) for application to reciprocal associations between a CS and US (and between one CS and another). It also elaborates upon their simplifying assumption about the mapping between associative strength and conditioned responding to provide an analysis of the origin and distribution of different conditioned responses. One benefit of these changes is that HeiDI cannot only account for phenomena upon which the reputation of the Rescorla and Wagner (1972) model was founded, but it also accounts for a broad range of other phenomena that have proven resistant to coherent theoretical analysis. HeiDI also makes novel predictions, including those examined here.
Declaration of competing interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability
OSF links are included in the manuscript.

Acknowledgements
The research was supported by the BBSRC (UK; BB/T004339/1; PI: RCH). We thank A. F. Iliescu for help with conducting the research, and Advanced Research Computing at Cardiff (ARCCA) for providing the computing resources necessary for the statistical analyses presented here.

References
Beckmann, J. S., & Chow, J. J. (2015). Isolating the incentive salience of reward-interests or personal relationships that could have appeared to influence the work reported in this paper.

Neurobiology of Learning and Memory 211 (2024) 107915
V. Navarro et al.

Konorski, J. (1948). Conditioned reflexes and neuron organization. Cambridge University Press,


Navarro, V., & Calmar: Canonical Learning Models and their Representations (0.6.1). https://cran.r-project.org/web/packages/calmar/index.html

Data curation, Conceptualization.


