Emergence of complex dynamics of choice due to repeated exposures to extinction learning

José R. Donoso[1], Julian Packheiser[2], Roland Pusch[2], Zhiyin Lederer[1], Thomas Walther[1], Metin Uengoer[3], Harald Lachnit[3], Onur Güntürkün[2] and Sen Cheng[1]*

[1]Institute for Neural Computation, Ruhr-Universität Bochum, Universitätstr. 150, 44801 Bochum, Germany; [2]Department of Biopsychology, Ruhr-Universität Bochum, Universitätstr. 150, 44801 Bochum, Germany; [3]Department of Psychology, Philipps-Universität Marburg, Gutenbergstraße 18, 35032 Marburg, Germany.

* Correspondence to: sen.cheng@rub.de

Abstract

Extinction learning, the process of ceasing an acquired behavior in response to altered reinforcement contingencies, is not only essential for survival in a changing environment, but also plays a fundamental role in the treatment of pathological behaviors. During therapy and other forms of training involving extinction, subjects are typically exposed to several sessions with a similar structure. The effects of this repeated exposure are not well understood. Here, we studied the behavior of pigeons across several sessions of a discrimination learning task in context A, extinction in context B, and a return to context A to test the context-dependent return of the learned responses (ABA renewal). By focusing on individual learning curves across animals, we uncovered a session-dependent variability of behavior: (1) During extinction, pigeons preferred the unrewarded alternative choice in one-third of the sessions, predominantly during the first one. (2) In later sessions, abrupt transitions of behavior at the onset of context B emerged, and (3) the renewal effect decayed as sessions progressed. We show that the observed results can be parsimoniously accounted for by a computational model based only on associative learning between stimuli and actions. Our work thus demonstrates the critical importance of studying the trial-by-trial dynamics of learning in individual sessions, and the power of “simple” associative learning processes.

Keywords: trial-by-trial dynamics, operant conditioning, pigeons, ABA renewal, decision-making

Author Contributions: Designed research: JRD, JP, RP, MU, HL, OG, SC; Performed research: JRD, JP, RP; Contributed new reagents or analytic tools: ZL, TW; Analyzed data: JRD, ZL, SC; Wrote the paper: JRD, JP, RP, MU, HL, OG, SC.

Conflict of Interest: The authors declare that they have no conflicts of interest.

Introduction

Animals modify their behavioral repertoire based on the consequences of their own actions. Whether a certain behavior is reinforced or not, respectively increases or decreases the likelihood that this behavior is repeated in a similar situation (Skinner 1938). This process of operant conditioning is not only pivotal for the discovery of purposeful actions, but also plays a fundamental role in the development of pathological behaviors such as drug addiction, overeating and gambling (Bouton 2011; Everitt and Robbins 2005; Hyman 2005; Kelley 2004; Redish 2004). These pathological behaviors might arise from reinforcement for operant actions such as approach, handling and consumption, which are under the control of natural stimuli such as, for example food, drug, games, and money, and context such as, for instance a restaurant, drug den or casino. Once a behavior is acquired, the ability to extinguish it as a result of altered reinforcement contingencies is also essential for survival. This so-called extinction learning is defined as the reduction of conditioned responses to a previously reinforced stimulus when that stimulus no longer signals reinforcement. The importance of extinction learning is emphasized by the fact that all vertebrate and invertebrate species tested exhibit this ability (Barad et al. 2006; Eisenberg and Dudai 2004; Galatzer-Levy et al. 2013; Gao et al. 2018; Gottfried and Dolan 2004; Lengersdorf et al. 2014; Milad et al. 2007; Stollhoff et al. 2005). Extinction learning is also relevant during therapy and other learning settings, where, for example, patients are trained to withdraw undesired behaviors, or when children are required to suppress recurrent spelling mistakes.

Although extinction may involve some erasure of the previously acquired memory (McClelland and Rumelhart 1985; Rescorla and Wagner 1972), there is strong evidence that it also involves new learning (Bouton 2019; Todd 2013; Trask et al. 2017). According to the latter view, during extinction, the previously acquired memory trace responsible for a particular behavior is inhibited by a secondary memory trace. This new learning seems to depend on the context for expression, which is compatible with the idea that contextual cues can support memory retrieval (Tulving and Thomson 1973). A prominent phenomenon in support of extinction as new learning is the renewal effect, where an extinguished behavior reemerges when subjects are removed from the context where extinction learning took place (Bernal-Gamboa et al. 2017; Bouton 2019; Nieto et al. 2017, 2020; Todd 2013; Todd et al. 2014). This context-dependence of extinction has a severe downside under conditions like exposure therapy, where a seemingly extinguished pathological behavior resurfaces when patients switch from a therapy context to their regular environment (Conklin and Tiffany 2002). Therefore, understanding extinction and the factors that influence the reappearance of the extinguished behavior can be helpful in developing treatments for several pathological behaviors and in preventing their reappearance (Bouton 2011; Conklin and Tiffany 2002).

Another relevant aspect of extinction learning is that it is not limited to a decrease in previously reinforced responses, but it can also drive the emergence of new, previously non-reinforced behaviors (Antonitis 1951; Fuller 1949; Grow et al. 2008; Lattal and Lattal 2012; Tinsley et al. 2002). This side-effect of extinction has been exploited to shape behavior. For example, when the extinction of problematic behaviors drives the emergence of socially acceptable responses, the latter can be reinforced in order to replace undesired behaviors with desired ones (Grow et al. 2008). In the same way, extinction-induced variability becomes particularly relevant in real world settings and experimental designs where there are multiple alternative choices in a particular situation. While there are studies of extinction in the presence of alternative responses (e.g., Winterbauer and Bouton 2010), research has focused mostly on the cessation of target responses, often neglecting the effect that extinction might have on other available actions (André et al. 2015; Méndez-Couz et al. 2019).

Experiments involving extinction learning, and/or testing for the renewal effect, rarely involve more than one conditioning-extinction sequence (e.g., Bernal-Gamboa et al. 2017; Bouton 2019; Nieto et al. 2017, 2020; Todd et al. 2014; Trask et al. 2017); but see (Anger and Anger 1976; Bullock 1960; Jenkins 1961). Outside the laboratory, however, an animal's learning history comprises multiple instances in which a period of conditioning is followed by extinction. Such multiple

conditioning-extinction sequences could differ in the stimuli involved, but may share a similar structure. Nevertheless, in most studies involving repeated extinction, the same association that is learned and extinguished in a given session is relearned and re-extinguished in subsequent sessions (Bai and Podlesnik 2017; Craig and Shahan 2016; Craig et al. 2019; Podlesnik and Sanabria 2011). In spite of its relevance, the effect of this repeated exposure to the same task structure with varying learning content has not been studied systematically. Another aspect that remains to be elucidated is the effect of the learning history both on the emergence of alternative behaviors, and on the prevalence of the renewal effect across sessions.

In the present study, we analyzed the choice behavior of pigeons undergoing multiple sessions of a discrimination learning task (Fig. 1). Each session consisted of three subsequent stages: (1) learning of correct responses to two novel visual stimuli in acquisition context A; (2) extinction learning for one of the learned stimulus-response mappings in context B, and (3) return to context A to test for ABA renewal of the extinguished mapping (e.g., Nieto et al. 2017). We used pigeons due to their ability to withstand multiple sessions of learning of more than 1000 trials per day. Moreover, their slow pace of learning allowed us to focus on the trial-by-trial dynamics of choice behavior in single sessions and individual subjects. This approach uncovered a rich repertoire of choice behavior associated with the extinction and renewal-test phases that are key to understanding associative learning mechanisms, and that are seldom reported. A parsimonious model consisting of an associative network and a winner-takes-all decision making process, where the action unit with the highest activation drives behavior could account for the rich behavioral phenomena observed in the data.

Materials and Methods

The behavioral data reported here were obtained in the same experiment in which electrophysiological data were collected to study the neural correlates of the reward prediction error in extinction learning and renewal (Packheiser et al. 2020a). The present study is independent from the previous study regarding focus, hypotheses, data analysis and interpretation. Specifically, the previous study did not analyze learning curves in individual sessions or across repeated extinction sessions. Data collection for the present study started after surgical procedures were completed.

Subjects. Eight experimentally naive, adult pigeons (*Columba livia*) of unknown sex and age obtained from private breeders were used as subjects in the present experiment. Birds were housed in individual wire-mesh cages or local aviaries within a colony room. The housing facilities were controlled for light cycles (12 h light/dark cycles starting at 8 am), temperature and humidity. All animals had *ad libitum* access to water and were kept between 80% and 90% of the free-feeding body weight. The food deprivation was necessary to keep the animals engaged in the experimental procedures. All animals were treated in accordance with the German guidelines for the care and use of animals in science. The experimental procedures were approved by a national ethics committee of the State of North Rhine-Westphalia, Germany and were in agreement with the European Communities Council Directive 86/609/EEC concerning the care and use of animals for experimental purposes.

Apparatus. The experimental procedures were conducted in custom-made Skinner boxes (35cm x 35cm x 35cm situated in sound-attenuating cubicles (80cm x 80cm x 80cm)(Packheiser et al. 2020b). Each Skinner box featured three rectangular pecking areas (5cm x 5cm) that were horizontally arranged on the rear wall. Depending on the type of Skinner box, either touch screens or translucent response keys combined with a mounted LCD flat screen monitor were used to track pecking responses. A feeder was located below the central pecking site (feeding area: ~1cm x 1cm) to deliver food rewards during the experiments. White LED strips mounted to the ceiling were used to illuminate the experimental chamber. Furthermore, red and green LED strips were attached to the ceiling to enable flexible contextual changes during the paradigm. If the animals successfully pecked onto a response key, an auditory feedback sound was presented. The hardware was controlled by a custom written MATLAB program (The Mathworks, Natick, MA, USA) using the

Biopsychology toolbox (Rose et al. 2008).

Procedure. Pigeons underwent 6 to 26 sessions of a discrimination learning task in which animals were subject to acquisition, extinction and a renewal-test phase within one session (Fig. 1A) (Packheiser et al. 2019; Starosta et. al. 2014). In each trial, animals were presented one visual stimulus. In response, the animal had to make either a left or a right choice at the end of the trial depending on the stimulus identity. During a session, the animals were confronted with four different stimuli presented in a pseudorandomized order. Two of the stimuli were associated with a left choice and the other two stimuli were associated with a right choice. Animals were pre-trained on two of the stimuli prior to the experimental sessions studied here. The precise nature of the pre-training procedure is explained in detail in (Packheiser et al. 2020a). Hence, two of these stimuli were familiar to the animals and served as control stimuli as well as fix points during the experiment. The other two stimuli were session-unique and the stimulus-response mapping had to be learned in the acquisition phase through trial-and-error. For the familiar stimuli, we used black and white stimuli featuring either a squared or a striped texture (see Figure 1A). Novel stimuli were always colored and featured different forms or shapes to allow for a discrimination both in color and spatial features.

Acquisition

Figure 1B shows a sequence of screen presentations during acquisition. In brief, each trial started with the presentation of an initialization key for up to 6s. A successfully registered key peck to the center response key triggered the sample presentation. One of four stimuli (see below) was presented for 2.5s on the center key. Following the stimulus presentation, the animals were required to confirm that they attended the target stimulus by pecking on the center key once more. After pecking on the confirmation key, the center key stimulus disappeared and the two choice keys were illuminated (a blue one on the left and a yellow one on the right, counterbalanced across animals). The animal had to decide on a left or a right choice depending on the identity of the stimulus that was presented earlier. If the animals made the correct choice (henceforth 'learned choice'), a 2s long reward period commenced during which the food hopper was illuminated, a sound occurred and food was available. In the case of an incorrect choice (henceforth 'alternative choice'), a different sound occurred and the lights in the chamber were turned off for 2s as a mild punishment. Consecutive trials were separated by an inter-trial-interval (ITI) of 4s duration. A detailed description of the events and transitions between periods within a single trial is depicted as a flowchart in Figure 1D. The acquisition phase comprised a minimum of 150 trials and ended once the animals satisfied all of the following criteria: the animals initialized 85% of the trials correctly, performed above 85% correctly in response to the novel stimuli and above 80% correctly is response to the two familiar stimuli. Performance values were calculated as a running average over the past 100 trials.

Extinction

The subsequent extinction phase was marked by two key differences as compared to the acquisition phase (Fig. 1C). (1) one of the novel stimuli was randomly chosen as the extinction stimulus, i.e., it was no longer followed by reward nor by punishment after any choice the animal made. Instead, the feedback phase was replaced by a 2s-long period void of feedback. (2) After the initialization of the trial by the animal, a red LED light (indicator of context B1) replaced the white house light used in the acquisition phase. The red LED light remained on until the end of the trial or until a punishment condition was met. To ensure that the physical identity of the red light was not driving behavioral effects in the extinction context, we also used a green LED light as an indicator for the extinction context (context B2). During extinction training, both the red and green context were present in each session and specifically associated with two experimental stimuli, namely one familiar and one novel stimulus for each context. Therefore, one of the contexts was always the context associated with extinction learning whereas the other context was not associated with extinction learning. Additionally, we also changed the audio cues for the correct

and incorrect responses to add another modality change highlighting the contextual switch. The extinction phase comprised a minimum of 150 trials and ended when the following conditions were all met: the animals initialized 85% of the trials correctly, performed above 80% correctly in response to the novel non-extinction stimulus and more than 75% correctly in response to the two familiar control stimuli, and emitted the learned choice in response to the extinction stimulus less than 20% of the time. All performance values were calculated as a running average over the past 100 trials.

Renewal-test

Finally, the renewal-test phase was used to study the return of the learned choice when the context was switched back to the acquisition context A (Fig. 1A). Importantly, the extinction stimulus remained without feedback to measure the renewal effect. The renewal-test phase lasted for a fixed number of 250 trials and required no behavioral criterion to end. Its end also marked the end of the session. Data acquisition was ended when the animals no longer provided any neural data for the study by Packheiser et al. (2020a).

Analysis. To quantify the preference for a target choice in a given phase, we counted the corresponding responses (*k*), and expressed them as a proportion of the total number of trials (*N*) in that phase (response rate = k/N). To assess the significance of a response count *k*, we calculated the probability *p(k)* of obtaining at least *k* responses by chance in *N* random trials under the null hypothesis. Since responses to the extinction stimulus are not rewarded during the extinction and renewal-test phases, our null-hypothesis assumes unbiased random responses, such that each one of the three possible outcomes (learned choice, alternative choice, or omission) can occur with probability ⅓. If the probability of observing at least *k* responses by chance was below a threshold of 0.05, we regarded the count as significant. This method, however, overlooks those cases where a non-significant number of responses are arranged in a chain of persistent responses, which is also unlikely to occur by chance. To consider those cases, we also measured the length *L* of the longest chain of persistent responses found on a specific phase (AP in Figs. 3C and 6A), and calculated the probability *p(L)* of obtaining a chain of at least *L* trials by chance in *N* random trials. Finally, the choice behavior was regarded as significant, if one of the aforementioned tests yielded a *p*-value below 0.05. We also used this measure to quantify the prevalence of a target response, which was defined as the proportion of animals expressing a significant number of target responses in a given session. To quantify the dependency of these response measures on the session number, we calculated the rank order correlation (Spearman's r) between session number and both preference and prevalence. Since the number of animals contributing with data gradually decreased with session number, we grouped the data in blocks encompassing a comparable number of data points: Sessions 12 to 23 in blocks of two sessions (9 to 12 data points per block), and the data from sessions 24 to 26 in one single block (9 data points). The rank-order correlations (Spearman's r) between the session number and the prevalence/prevalence of a particular response across animals was obtained from the grouped data unless stated otherwise. The p-values accompanying the reported correlations correspond to a t-test for the significance of the respective correlation coefficient (r).

To visualize the behavior in response to a specific stimulus within single sessions, we plotted the cumulative record of responses to that stimulus as a function of trial number (Fig. 3). For each trial, the choice behavior of the animal, namely, omissions, alternative choice, and learned choice, were encoded as 0, -1, and 1, respectively. To assess the presence of abrupt transitions of behavior upon the onset of the extinction phase, we focused on the responses to the first 5 presentations of the extinction stimulus under context B. Our null-hypothesis (i.e., no behavioral change) assumes that animals continue to emit the learned choice with the 85% probability required to accomplish the acquisition phase. Thus, if animals emit at least 3 non-reinforced choices (alternative choices or omissions) within the first 5 trials of the extinction phase, the null-hypothesis is rejected ($p = 0.022$, Binomial test), and the behavioral transition is considered abrupt. Otherwise, the behavioral transition is considered smooth.

Associative network and decision making model. The model consists of a simple network that associates sensory input with two motor outputs (triangles), one each for the left (L) and right (R) responses (Fig. 5). Binary sensory units (ovals) signal the presence or absence of a specific stimulus (including the context) with a 1 or 0, respectively. These sensory units provide excitatory and inhibitory input to the motor units. Hence, the total synaptic input to the motor units is given by:

$$u = W_{exc}s - W_{inh}s,$$

where u is a two element vector containing the input to the L and R motor units, W_{exc} and W_{inh} are matrices containing the excitatory and inhibitory synaptic weights, respectively, and $s$is a binary vector specifying the set of stimuli that are present in a given trial. The motor units are rectifying linear units, i.e. they have a threshold at 0, which are driven by the net synaptic input and excitatory noise. Hence, the activation of the motor units is given by:

$$m = ReLU(u + \varepsilon),$$

where m is a two element vector describing the activation of L and R, and ε is a two element vector containing the noise inputs to L and R. These are drawn from two independent uniform distributions on the interval (0, 1). The behavioral choice corresponds to the motor unit with the highest activation in the presence of a given stimulus and context:

$$choice = argmax(m).$$

If the total input (synaptic plus noise) to both motor units are equal or lower than 0, no response is selected, resulting in a choice omission.

If a reward is delivered upon responding, excitatory connections between the active sensory units and the responding motor unit are reinforced. Otherwise, excitatory connections remain unchanged. Conversely, when a reward is not delivered, inhibitory connections between the active sensory units and the responding motor unit are reinforced. Otherwise, inhibitory connections remain unchanged. The value of the synapses (i.e. associative strengths) are updated according to:

$$\Delta w_{ij} = \lambda_{exc,inh}(w^{\infty}_{exc,inh} - w_{ij}),$$

where $\Delta w_{ij}$is the increase of the synapse connecting input *i* with motor unit *j*, $\lambda_{exc,inh}$corresponds to the learning rate of excitation or inhibition, and $w^{\infty}_{exc,inh}$is the maximum possible value that excitatory or inhibitory weights can reach, i.e., their respective saturation values (Bush and Mosteller 1951; Mackintosh 1975; Pearce and Hall 1980; Rescorla and Wagner 1972; Schmajuk and Moore 1985).

For simplicity, all the synapses in the pigeon-models had the same parameters: The learning rates of all the excitatory and inhibitory connections to motor units (λ_e and λ_i, respectively) were set to 0.02. Except for the results shown in Figure 5 where λ_i was set either to 0.005, 0.01 or 0.02. This was a way to test the effect of three different ratios of inhibitory/excitatory learning rates (λ_i/λ_e in Fig. 5B and C). All synaptic weights saturated at a value of 20.

Results

Session-to-session variability of choice behavior. In the following, we focus on the expression of alternative choices (one for which the pigeon was not rewarded in the acquisition phase) during the extinction phase, and on the return of the learned choice during the renewal-test phase. In particular, we analyze the overall expression of these responses, as well as their dependence on the number of sessions an animal was exposed to extinction (Fig. 2). Figure 2A1 shows the preference for the learned choice in response to the extinction stimulus as a function of session number. During the acquisition phase (blue trace), animals exhibited a strong preference for the learned choice, which was not significantly correlated with the session number ($r = -0.27$; $p = 0.29$). This indicates that animals successfully learned the correct mapping between the novel stimulus and the pecking side repeatedly. During the 2nd half of the extinction phase (red trace), the low preference for the learned choice shows that animals successfully extinguished the learned association. Again, this behavior was also not significantly correlated with session number ($r = -0.40$; $p=0.098$). It is noteworthy that extinction of one novel stimulus-response mapping did not interfere with the learned responses to the three control stimuli in interspersed trials since performance remained high near or above 90% for the control stimuli (Fig. 2B). This also indicates that the context change to B alone did not drive the change in behavior in response to the extinction stimulus. Upon return to context A, the preference for the learned choice reemerged when the extinction stimulus was presented (black trace), which constitutes the ABA renewal effect. In the first session, the proportion of learned choice at the beginning of the renewal testing phase was significantly higher than in the second half of extinction ($p<0.05$, z-test). Similarly, the learned responding increased for two of the three control stimuli significantly as well ($p<0.05$, z-test) indicating a generalized effect of context change from B to A, but the increase was quite small (Fig. 2B).

Interestingly, renewal in response to the extinction stimulus changed across repeated sessions of the ABA-renewal paradigm. Across animals, both the average preference for the learned choice (Fig. 2A1, black trace), and the prevalence of these responses (Fig. 2A1, gray bars) were strongly correlated with session number ($r = -0.78$; $p = 0.00024$ and $r = -0.76$; $p = 0.00028$, respectively). This session correlation was not significant in the responses to the three control stimuli ($p>0.05$, Fig. 2B). Nevertheless, the preference for the learned choice in response to the control novel stimulus (black trace in Fig. 2B, left) was slightly but significantly higher than during the second half of extinction in 14 out of the 18 session blocks ($p<0.05$; z-test; stars in Fig. 2B, left). The decay of the renewal effect observed across animals was also apparent in individual pigeons (Fig. 2A1, middle). Four animals showed a significant negative correlation between preference and session number (Fig. S1B). Furthermore, the largest response rates for the learned choice during renewal-test occurred within the first four sessions for every individual pigeon (warm colors in Fig. 2A1, middle).

Figure 2A2 shows the preference for the alternative choice in response to the extinction stimulus as a function of session number. During the extinction phase (red trace), the preference for the alternative choice across animals was significantly higher ($p<0.05$; z-test; stars in Fig. 2A2) than during the second half of the acquisition phase (blue trace) for every session block. This difference was not mirrored by the responses to the control stimuli (Fig. 2B). Furthermore, one-third of the individual sessions exhibited a significant preference for the alternative choice during extinction (Fig. 2A2, bottom). Remarkably, most of these sessions (90%) exhibited chains of 5 to 25 consecutive trials (median: 7 trials) where the animals opted persistently for the alternative choice, further indicating that these responses are not random occurrences. We refer to the length of these chains as 'alternative persistence' (AP in Fig. 3C). However, the preference for the alternative choice during extinction was not significantly correlated with session number ($r= -0.44$; $p = 0.070$), except for two individual animals (Fig. S1A). In spite of this weak correlation to session number, we nevertheless observed that the prevalence of the alternative choice was not evenly distributed across sessions (Fig. 2A2, gray bars): A significantly larger proportion of animals expressed this type of responses during the first session (76%), when compared to the proportion found in all the

remaining sessions combined (32%; $p = 0.0062$; z-test). To confirm this asymmetry, we also compared every other session to all the remaining sessions combined, and found no significant difference in the proportions ($p > 0.17$; z-test).

Visualizing choice-behavior in single sessions. To study the session-to-session changes in choice behavior in more detail, we visualized the entire time-course of learning within individual sessions. We did this by focusing on the cumulative record of successive behavioral responses as a function of trial number (Fig. 3; Gallistel et al. 2004; Leslie et al. 2005). However, we departed from the traditional 'unipolar' encoding of the responses, which focuses only on the learned choice by signaling its presence or absence with 1 or 0, respectively. Instead, we encoded each learned choice, alternative choice, or omission as +1, -1 and 0, respectively. Under this bipolar encoding, the slope in the cumulative curve reveals biases towards specific responses (Fig. 3A, black trace): A positive or negative slope indicates a tendency to prefer the learned choice or the alternative choice, respectively. A slope of 0 indicates either a continuous chain of omissions or random mixtures of different responses. In Figure 3A, for example, the bipolar encoding (black trace) reveals that the extinction phase (gray area) is characterized by a persistent selection of the alternative choice. For comparison, the gray traces in Figure 3A and B illustrate how the unipolar encoding conceals the choice behavior. By focusing on a target response, the unipolar encoding can create the impression that the behavior observed in a grand average across sessions is representative of single sessions (Fig. 3B). Furthermore, the average unipolar learning curve looks similar to what one expects to find in ABA renewal experiments that use very different experimental designs from ours (e.g. Bouton et al. 2011; Gottfried and Dolan 2004; Lissek et al. 2015). Since our focus is on choice behavior, in the following, we use bipolar learning curves to study the variability of behavior across multiple sessions (Fig. 3).

Diversity of learning curves across sessions. From session to session, individual animals exhibited a rich repertoire of learning curves. As an example, Figure 3C shows eight sessions obtained from the same animal. During the extinction phase (gray area), behavior was often dominated either by random responses (e.g., sessions 2, 7 and 15) or omissions (e.g., sessions 20 and 24). In some cases, however, the learning curves developed negative slopes (e.g., in sessions 1, 6 and 14), revealing the preference for the alternative choice (AC in Fig. 3C) described in the previous analysis (Fig. 2A2). In the same way, the renewal-test phase was characterized by a variable re-emergence of the learned response (stars on top right of panels; see also Fig. 2A1, bottom). Upon return to context A, animals emitted a few learned responses before re-extinction took place (e.g., sessions 1, 7 and 20), or renewal was not present at all (e.g., sessions 15 and 24; see also Fig. 2A1, bottom). In some cases, animals persisted on the learned choice for most of the renewal-test phase (e.g., sessions 2 and 6; see also Fig. 2A1, middle).

Another puzzling aspect uncovered by the bipolar encoding were the abrupt transitions upon the onset of the extinction phase. According to a widespread view of extinction learning, as animals experience a withdrawal of reinforcers upon the onset of context B, they initially persist on the previously reinforced choice for several trials before gradually changing their behavior towards omissions (Nevin 2012; Podlesnik and Shahan 2009, 2010). Indeed, several learning curves in our data were characterized by these smooth transitions (e.g., sessions 1 and 2 in Fig. 3C). In some sessions, however, the change in behavior occurred abruptly upon the onset of context B (e.g., sessions 6, 7, 20 and 24 in Fig. 3C). These changes were apparently driven by the switch from the acquisition to the extinction context. To assess the presence of this peculiar behavior, we considered a transition abrupt if a pigeon emitted at least 3 non-reinforced choices (alternative choices or omissions) within the first 5 trials of the extinction phase ($p = 0.022$, Binomial test). According to this analysis, 39% of the sessions exhibited such abrupt transitions.

Four extinction types and their session to session variability. So far, we have identified two remarkable features of extinction learning, namely, the preference for the alternative choice, and abrupt transitions of behavior upon the onset of the extinction phase. Based on the presence or absence of these features, we arranged the learning curves in four classes, and then assessed the

prevalence of these classes across pigeons and sessions (Fig. 4). Figure 4A shows a two-by-two matrix of panels where the left and right columns correspond to smooth and abrupt transitions, respectively. And the top and bottom rows to the absence and presence of alternative choices, respectively. Each panel contains an overlay of all the learning curves (gray traces) belonging to the respective class, and their average trace (colored line). Thus, class 1 corresponds to the canonical learning curves, which accounted for 44% of the total sessions (Fig. 4A1). The majority of the learning curves (56%) deviated from the canonical one in at least one major aspect: either because animals exhibited abrupt transitions at the onset of context B (class 2), favored the alternative choice over omissions during extinction (class 3), or both (class 4, Fig. 4A4). Next, we analyzed the way in which the learning curve types described in Figure 4A were distributed across pigeons and sessions (Fig. 4B and C, respectively). Individually, pigeons did not consistently express a particular class of learning curve. In fact, all pigeons, except for one (for which only 6 sessions were available), exhibited all four learning curves classes (Fig. 4B). To get insights on the effects of re-testing the animals repeatedly**,** we also analyzed the session-to-session changes in the distribution of learning curve classes (Fig. 4C). Curves with smooth transitions (classes 1 and 3) dominated the first session, and occurred less frequently in later sessions. To confirm this observation, we used the method proposed by Gallistel et al. (2004) to determine the change point of the learning curves during extinction. We found that the change of behavior occurred significantly later in the first session as compared to the subsequent sessions ($p = 0.023$; KS-Test, Fig. S2).

Can associative learning account for the observed behavioral complexity? So far, we have reported three key findings in our behavioral data: (1) Smooth transitions of behavior at the onset of context B are more prevalent during the first exposure to the extinction task than in later sessions. (2) During extinction, pigeons express a preference for the alternative choice in nearly one-third of the sessions, most prominently in the first session. And (3), the renewal effect shows an overall decay as sessions progress. To study the mechanisms underlying these phenomena, we implemented a parsimonious model aimed at capturing the associative aspect of the task (Fig. 5). This model embodies two fundamental principles of associative learning. First, representations of the present stimuli can freely establish both excitatory and inhibitory associations with motor actions. Second, these associations are modulated by reinforcement contingencies. Additionally, we treat the context as just another stimulus that can establish direct inhibitory and excitatory associations with specific motor actions, in accordance with recent studies (Bernal-Gamboa et al. 2018; Bouton 2019; Nieto et al. 2017; Todd 2013; Todd et al. 2014). This is further supported by the fact that context changes do not necessarily imply that a complex environment being present for a long period of time is totally changed. Instead, it is sufficient, that a salient fraction of contextual stimuli is present during the presentation of the nominal stimulus (for an explicit proof, see Lachnit 1986; Üngör and Lachnit 2006; for successful application, see Uengoer et al. 2013).

Figure 5 summarizes the components of the model, which operate as follows: Sensory units (ovals) signal the representation of the context and specific stimuli in working memory. These units can establish direct excitatory connections with the motor units (triangles) mediating the left (L) and right (R) responses. They can also inhibit the motor units via interneurons (circles). Thus, excitatory and inhibitory associations between stimuli and actions are mediated by two independent pathways, as previously suggested (Felsenberg et al. 2018; Ghazizadeh et al. 2012). The decision making is performed by a simple winner-takes-all mechanism, where the motor unit with the highest activation drives the corresponding behavioral response. If both motor units are inhibited below their threshold of activation, no response is selected, resulting in a choice omission. Excitatory connections to motor units are reinforced if a reward is delivered, and remain unchanged otherwise. Conversely, connections onto interneurons are reinforced only if a reward is not delivered, and remain unchanged otherwise. The synaptic weights (i.e., associative strengths) mediating these connections grow asymptotically towards a saturation value, in accordance with standard models of associative learning (Bush and Mosteller 1951; Mackintosh 1975; Rescorla and Wagner 1972).

Associative learning accounts for the features observed in the behavioral data. We subjected a population of 20 pigeon-models to a sequence of 20 training sessions where the extinction

context and the extinction stimulus were selected randomly, as in the behavioral experiments. For simplicity, all the synapses in the pigeon-models had the same parameters (see Methods for details). The only parameter that was varied across simulations was the learning rate of inhibition (λ_i), which allowed us to test the effect of three different ratios of inhibitory/excitatory learning rates (λ_i/λ_e in Fig. 6B and C). Figure 6A shows a sample of four learning curves obtained from one of the pigeon-models. These curves exhibit the three features we uncovered in the pigeon behavior: preference for alternative choices during extinction (sessions 1 and 4), abrupt transitions of behavior upon onset of context B (session 4), and absence of the renewal effect (sessions 7 and 8). Like the pigeons in our study, the model shows a preference for the alternative choice during extinction, which rapidly declines as a function of session number (Fig. 6B, top). The proportion of pigeon-models emitting a significant number of alternative responses (Fig. 6B, bottom) resembles the findings in our experimental data (see Fig. 2A). Across the set of parameters tested, the significant expression of alternative choices was limited to the first few sessions. Also qualitatively reproducing our observations in pigeons, the expression of the renewal effect declined as a function of session number, as evidenced by both the average preference for the learned choices during the renewal-test phase (Fig. 6C, top), and in the proportion of pigeon-models emitting a significant number of learned choices (Fig. 6C, bottom). Since no parameter was adjusted between sessions, the variability in behavior stemmed solely from the history of learning. In particular, due to the associations between the context and specific responses (Bouton 2019; Leitenberg et al. 1970; Todd 2013; Todd et al. 2014; Winterbauer and Bouton 2010), which are carried over from session to session. The precise mechanisms underlying this variability are explained in detail in the supplementary information (see Fig. S3 and accompanying text). Briefly, during extinction, negative associations between context B and the learned choice, and positive associations between context B and the contralateral response act synergistically to tilt the balance in favor of the alternative choice. As sessions progress, the associative strengths maintaining this imbalance saturate quickly, favoring the expression of alternative choices during the first session. The prevalence of smooth transitions during the first session can be explained as follows: Upon the first exposure to the extinction phase, negative associations between the context and specific motor responses require several trials to build up (e.g., Fig 3SC1), but once established, they can exert an effect on the behavior in later sessions (e.g., Fig. 3SC2-4). This inheritance of context-response associations can lead to abrupt transitions of behavior upon switching to the extinction context, which can only occur in sessions after the first exposure to a context-dependent extinction phase. The renewal effect, on the other hand, occurs due to the release of the inhibition exerted by context B on a specific response (Bouton 2019; Todd 2013; Todd et al. 2014; Trask et al. 2017). After many sessions of training, these negative associations between context and responses reach their saturation value. As a consequence, the learned response can no longer be rescued by the release of context-inhibition. Thus, the decay of the renewal effect with sessions is a natural consequence of the existence of an asymptote of conditioning; a property that is ubiquitous across models of associative learning (Bush and Mosteller 1951; Mackintosh 1975; Pearce and Hall 1980; Rescorla and Wagner 1972; Schmajuk and Moore 1985).

Discussion

We have analyzed the behavior of pigeons subject to multiple sessions of a discrimination learning task in context A, extinction in context B, and a return to context A to test for the return of the learned response. By focusing on learning curves from individual animals and single sessions, we uncovered a rich diversity and dynamics of behavior during the extinction and renewal-test phases: (1) Upon the onset of the extinction phase, pigeons tended to persist on the learned choice mostly during the first session, whereas abrupt transitions of behavior emerged exclusively in later sessions. (2) During extinction, pigeons preferred the unrewarded alternative choice in one third of the sessions, predominantly during the first one. And (3), the renewal effect was intermittent and decayed as sessions progressed. The present results have specific implications for the treatment of pathological behaviors (Podlesnik et al. 2017), for a review). The higher prevalence of alternative choices during the first session in the present experiment indicates that treatments aimed at

extinguishing problem behavior and promoting alternative, appropriate behavior should make particular use of the early stage of extinction treatment to reinforce the more desirable alternative responses that might emerge (Antonitis 1951; Fuller 1949; Grow et al. 2008; Lattal and Lattal 2012; Tinsley et al. 2002). Moreover, the decay of the renewal effect with session number suggests that repetitions of a treatment involving extinction can be helpful in order to prevent the context-dependent reemergence of undesired behaviors. To reveal potential mechanisms of this behavioral variability, we used a computational model to show that associative learning, in combination with a winner-takes-all decision process, can express a variability of behavior similar to that observed in the data.

The importance of studying the dynamics of learning. So far, research has mostly neglected the learning dynamics and variability of behavior, and instead focused on a few response types that were studied by population averages. This is particularly true for extinction learning, but also in other forms of learning such as, for example sensorimotor learning (Cheng and Sabes 2006, 2007). A further problem here is the tradition of pooling and averaging across subjects and sessions. Such analyses not only conceal the complexity of behavior during a learning task (André et al. 2015; Gershman et al. 2010; Larrauri and Schmajuk 2008; Lissek et al. 2015; Wirth et al. 2003), but also obscure the dynamics of learning within and across sessions (Todd 2013; Trask et al. 2017). For example, in previous studies, abrupt transitions might have been masked either because subjects were exposed to only one session (Larrauri and Schmajuk 2008; Redish et al. 2007), or because results were pooled across several sessions/subjects (Lissek et al. 2015), or they were not present because reinforcers were gradually reduced over time (Gershman et al. 2010). Another important aspect that can be lost in a grand average is the dynamics of the renewal effect across sessions. Indeed, in previous studies, the intermittency of the renewal effect as well as its overall decay with session number could have been masked by the pooling across sessions (Lissek et al. 2015). Furthermore, computational models usually assume that the properties of the average curve are representative of those of individual curves (Gershman et al. 2010; Larrauri and Schmajuk 2008; Redish et al. 2007). This assumption prevents models from giving a comprehensive account of the phenomenon of extinction learning, in particular with regard to the expression of choice behavior and its variability across subjects and sessions.

Preference for alternative choices and resurgence. The expression of alternative choices in our experiment could be related to the phenomenon of resurgence, wherein a previously reinforced and then extinguished response reappears during a period of extinction for a subsequently learned response (Leitenberg et al. 1970; Winterbauer and Bouton 2010). Thus, if we consider that an alternative choice in a given session might correspond to a response that has been reinforced and extinguished in a previous session, this previously extinguished response might "resurge" in the current session as an alternative choice once the reinforced choice is extinguished. However, in our results, alternative choices occur prominently already in the first extinction session, before any responses had been extinguished. In addition, resurgence cannot account for any of our other findings, namely, abrupt transitions, lack of renewal, and the dynamics of these behaviors.

Function and mechanisms of behavioral variability. The emergence of apparently purposeless behavior is puzzling. If one particular response is abandoned due to the lack of reinforcers, and an alternative is picked as a result, why would an animal persist on an unrewarded alternative over omissions? Evolutionary theories of foraging have proposed reasons why probing alternative behaviors might pay under reduced reward conditions (Anselme and Güntürkün 2019; Gharib et al. 2004), but these arguments explain the ultimate level, thereby lacking proximate, mechanistic explanations. From the perspective of reinforcement learning, explaining unrewarded alternative choices is difficult, as they do not provide any added value relative to an omission, and incurs the cost of the extra energy spent. Furthermore, models based on statistical inference predict that all the choices available in the presence of the extinction stimulus should be extinguished, and therefore, the extinction phase should be characterized by the absence of active choices (Redish et al. 2007). Our computational model, on the other hand, offers a parsimonious, mechanistic account by showing that associative learning, combined with a winner-takes-all decision making

process, predicts not only the emergence of variability across sessions, but also persistence on previously unrewarded responses.

Scope of the model. The purpose of our model was to test whether highly variable individual behaviors can emerge from simple associative learning. In particular, the law governing the update of associative strengths in our model does not consider an associability term (or salience), which previous models have used to describe the selective attention certain stimuli become due to their relative predictive power of desired/undesired outcomes (Bush and Mosteller 1951; Mackintosh 1975; Pearce and Hall 1980; Rescorla and Wagner 1972; Schmajuk and Moore 1985). In some of these models, this term is updated from trial to trial, thereby modulating the learning rates dynamically within single sessions (Mackintosh 1975; Pearce and Hall 1980; Schmajuk and Moore 1985) (for empirical evidence for learning-induced changes in selective attention, see Feldmann-Wüstefeld et al. 2015; Lucke et al. 2013; Koenig et al. 2017a, 2017b). Curiously, our model exhibits extinction learning at varying speeds (see Figs. 3B and 4), even though it lacks a term modulating the learning rates; a phenomenon that could be otherwise attributed to attentional variations due to reward expectancy, as suggested by previous models.

Associative learning: simple explanations for seemingly complex empirical observations. Although our model provides a parsimonious explanation of the observed behavior in terms of associative learning, it cannot rule out the involvement of higher-order cognitive functions. It could be argued, for example, that the decay of the renewal effect with session number (Fig. 2A1) reflects learning about the structure of the task (Livesey et al. 2019; Zentall et al. 2014): As sessions progress, animals might learn that a switch from context B back to context A does not predict a return of the reinforcement contingency of the acquisition phase. Hence, the renewal effect would decay gradually as animals experience more and more sessions with the same ABA structure. Since such putative abstract-rule learning is not perfect, some forgetting or attentional fluctuations might lead to the intermittent reappearance of the renewal effect (Fig. 2A1, bottom), but overall, renewal decays with experience. Along the same lines, the dominance of smooth transitions in the first session, and the appearance of abrupt transitions at later sessions (Fig. 4C) might also reflect some form of abstract-rule learning: In the first session, animals experience a withdrawal of reinforcers when the context is switched to B for the very first time. Not knowing that reward contingency is linked to context, the animals will initially persist on the previously reinforced response for several trials before gradually changing their behavior as a result of the absence of the expected reward. Such behavioral momentum might be reduced in later sessions as animals learn that a change from the acquisition context A to the extinction context B signals a change in the reward contingency. The application of this hypothetical rule might also be subject to attentional fluctuations and other sources of noise, resulting in the observed intermittent pattern of abrupt transitions. Yet another possibility to explain some of our results would be in terms of learned equivalence (Honey and Hall 1989; Zentall et al. 2014). In this framework, pigeons exposed to repeated sessions of our experimental paradigm could learn to group the stimuli under two context-sensitive categories, namely 'Peck-Left' and 'Peck-Right'. During extinction, alternative responses would emerge from a re-categorization of the extinction stimulus. After a few sessions, one could hypothesize that animals learn that the extinction context signals this re-categorization, giving rise to abrupt transitions of behavior upon the onset of context B. Based on our data, we cannot rule out that these hypothetical behavioral strategies drive the observed complex behaviors during extinction and renewal-test. However, in accordance with Morgan's Canon (Morgan 1903), we favor the parsimonious explanation of our results provided by our simple associative learning model over an interpretation in terms of higher-order cognition.

The role of context in extinction and renewal. A key aspect of our model is that the context can directly establish excitatory and inhibitory associations with specific responses (Bouton 2019; Rescorla 1993, 1997; Todd 2013; Todd et al. 2014). This feature was critical for our model's ability to account for the data. In particular, alternative responses emerged in the model, because it learned an excitatory association between the extinction context and the alternative choice either during an earlier phase or during the extinction phase when the response to the control stimuli are

rewarded. However, it has been suggested that context can also establish associations with the representations of the outcome (Baker et al. 1991; Pearce and Hall 1979), or modulate the association between response and outcome (occasion setting hypothesis) (Trask and Bouton 2014). Since these hypotheses are not mutually exclusive, our results cannot rule out a scenario where context exerts a direct or modulatory influence over the representation of the outcome. However, if the context was limited to that role, it is not clear how it could drive the emergence of alternative responses, or abrupt transitions of behavior upon the onset of the extinction phase in a purely associative learning scenario.

In conclusion, we have uncovered a rich variability of behavior in extinction learning and renewal that so far has remained concealed in population averages. Even though these complex behaviors appear to reflect abstract rule learning, we have demonstrated that associative learning can generate similarly complex behavior without resorting to higher-order cognitive processes.

Acknowledgments

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 316803389 – SFB 1280, projects F01, A01, A14 and A15.

Compliance with Ethical Standards

Funding: This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 316803389 – SFB 1280, projects F01, A01, A14 and A15.

Conflict of Interest: The authors declare that they have no conflicts of interest.

Ethical approval: All animals were treated in accordance with the German guidelines for the care and use of animals in science. The experimental procedures were approved by a national ethics committee of the State of North Rhine-Westphalia, Germany and were in agreement with the European Communities Council Directive 86/609/EEC concerning the care and use of animals for experimental purposes.

References

André MAE, Güntürkün O, Manahan-Vaughan D. 2015. The metabotropic glutamate receptor, mGlu5, is required for extinction learning that occurs in the absence of a context change. *Hippocampus* **25**: 149–158.

Anger D, Anger K. 1976. Behavior Changes During Repeated Eight-Day Extinctions. *J Exp Anal Behav* **26**: 181–190.

Anselme P, Güntürkün O. 2019. How foraging works: Uncertainty magnifies food-seeking motivation. *Behav Brain Sci* **42**: e35.

Antonitis JJ. 1951. Response variability in the white rat during conditioning, extinction, and reconditioning. *J Exp Psychol* **42**: 273–281.

Bai JYH, Podlesnik CA. 2017. No impact of repeated extinction exposures on operant responding maintained by different reinforcer rates. *Behav Processes* **138**: 29–33.

Baker AG, Steinwald H, Bouton ME. 1991. Contextual conditioning and reinstatement of extinguished instrumental responding. *Q J Exp Psychol Sect B* **43**: 199–218.

Barad M, Gean P-W, Lutz B. 2006. The Role of the Amygdala in the Extinction of Conditioned Fear. *Biol Psychiatry* **60**: 322–328.

Bernal-Gamboa R, Nieto J, Uengoer M. 2017. Effects of extinction in multiple contexts

on renewal of instrumental responses. *Behav Processes* **142**: 64–69.
Bernal-Gamboa R, Nieto J, Uengoer M. 2018. Removing but not adding elements of a context affects generalization of instrumental responses. *Learn Behav* **46**: 256–264.
Bouton ME. 2019. Extinction of instrumental (operant) learning: interference, varieties of context, and mechanisms of contextual control. *Psychopharmacology (Berl)* **236**: 7–19.
Bouton ME. 2011. Learning and the persistence of appetite: Extinction and the motivation to eat and overeat. *Physiol Behav* **103**: 51–58.
Bouton ME, Todd TP, Vurbic D, Winterbauer NE. 2011. Renewal after the extinction of free operant behavior. *Learn Behav* **39**: 57–67.
Bullock DH. 1960. Repeated Conditioning-Extinction Sessions as a Function of the Reinforcement Schedule. *J Exp Anal Behav* **3**: 241–243.
Bush RR, Mosteller F. 1951. A mathematical model for simple learning. *Psychol Rev* **58**: 313–323.
Cheng S, Sabes PN. 2007. Calibration of Visually Guided Reaching Is Driven by Error-Corrective Learning and Internal Dynamics. *J Neurophysiol* **97**: 3057–3069.
Cheng S, Sabes PN. 2006. Modeling sensorimotor learning with linear dynamical systems. *Neural Comput* **18**: 760–793.
Conklin CA, Tiffany ST. 2002. Applying extinction research and theory to cue-exposure addiction treatments. *Addiction* **97**: 155–167.
Craig AR, Shahan TA. 2016. Experience with dynamic reinforcement rates decreases resistance to extinction. *J Exp Anal Behav* **105**: 291–306.
Craig AR, Sweeney MM, Shahan TA. 2019. Behavioral momentum and resistance to extinction across repeated extinction tests. *J Exp Anal Behav* **112**: 290–309.
Eisenberg M, Dudai Y. 2004. Reconsolidation of fresh, remote, and extinguished fear memory in medaka: old fears don't die. *Eur J Neurosci* **20**: 3397–3403.
Everitt BJ, Robbins TW. 2005. Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nat Neurosci* **8**: 1481–1489.
Feldmann-Wüstefeld T, Uengoer M, Schubö A. 2015. You see what you have learned. Evidence for an interrelation of associative learning and visual selective attention. *Psychophysiology* **52**: 1483–1497.
Felsenberg J, Jacob PF, Walker T, Barnstedt O, Edmondson-Stait AJ, Pleijzier MW, Otto N, Schlegel P, Sharifi N, Perisse E, et al. 2018. Integration of Parallel Opposing Memories Underlies Memory Extinction. *Cell* **175**: 709-722.e15.
Fuller PR. 1949. Operant Conditioning of a Vegetative Human Organism. *Am J Psychol* **62**: 587–590.
Galatzer-Levy IR, Bonanno GA, Bush DE, LeDoux J. 2013. Heterogeneity in threat extinction learning: substantive and methodological considerations for identifying individual difference in response to stress. *Front Behav Neurosci* **7**: 55.
Gallistel CR, Fairhurst S, Balsam P. 2004. The learning curve: Implications of a quantitative analysis. *Proc Natl Acad Sci U S A* **101**: 13124–13131.
Gao M, Lengersdorf D, Stüttgen MC, Güntürkün O. 2018. NMDA receptors in the avian amygdala and the premotor arcopallium mediate distinct aspects of appetitive extinction learning. *Behav Brain Res* **343**: 71–82.
Gershman SJ, Blei DM, Niv Y. 2010. Context, learning, and extinction. *Psychol Rev* **117**: 197–209.
Gharib A, Gade C, Roberts S. 2004. Control of Variation by Reward Probability. *J Exp*

Psychol Anim Behav Process **30**: 271–282.
Ghazizadeh A, Ambroggi F, Odean N, Fields HL. 2012. Prefrontal Cortex Mediates Extinction of Responding by Two Distinct Neural Mechanisms in Accumbens Shell. *J Neurosci* **32**: 726–737.
Gottfried JA, Dolan RJ. 2004. Human orbitofrontal cortex mediates extinction learning while accessing conditioned representations of value. *Nat Neurosci* **7**: 1144–1152.
Grow LL, Kelley ME, Roane HS, Shillingsburg MA. 2008. Utility of Extinction-Induced Response Variability for the Selection of Mands. *J Appl Behav Anal* **41**: 15–24.
Honey RC, Hall G. 1989. Acquired equivalence and distinctiveness of cues. *J Exp Psychol Anim Behav Process* **15**: 338–346.
Hyman SE. 2005. Addiction: A Disease of Learning and Memory. *Am J Psychiatry* **162**: 1414–1422.
Jenkins HM. 1961. The effect of discrimination training on extinction. *J Exp Psychol* **61**: 111–121.
Kelley AE. 2004. Memory and Addiction: Shared Neural Circuitry and Molecular Mechanisms. *Neuron* **44**: 161–179.
Koenig S, Kadel H, Uengoer M, Schubö A, Lachnit H. 2017a. Reward draws the eye, uncertainty holds the eye: Associative learning modulates distractor interference in visual search. *Front Behav Neurosci* **11**: 128.
Koenig S, Uengoer M, Lachnit H. 2017b. Attentional Bias for Uncertain Cues of Shock in Human Fear Conditioning: Evidence for Attentional Learning Theory. *Front Hum Neurosci* **11**: 266.
Lachnit H. 1986. Transswitching and contextual conditioning. *Pavlov J Biol Sci* **21**: 160–172.
Larrauri JA, Schmajuk NA. 2008. Attentional, associative, and configural mechanisms in extinction. *Psychol Rev* **115**: 640–676.
Lattal MK, Lattal KA. 2012. Facets of Pavlovian and operant extinction. *Behav Process* **90**: 1–8.
Leitenberg H, Rawson RA, Bath K. 1970. Reinforcement of Competing Behavior during Extinction. *Science* **169**: 301–303.
Lengersdorf D, Stüttgen MC, Uengoer M, Güntürkün O. 2014. Transient inactivation of the pigeon hippocampus or the nidopallium caudolaterale during extinction learning impairs extinction retrieval in an appetitive conditioning paradigm. *Behav Brain Res* **265**: 93–100.
Leslie JC, Shaw D, Gregg G, McCormick N, Reynolds DS, Dawson GR. 2005. Effects of Reinforcement Schedule on Facilitation of Operant Extinction by Chlordiazepoxide. *J Exp Anal Behav* **84**: 327–338.
Lissek S, Glaubitz B, Güntürkün O, Tegenthoff M. 2015. Noradrenergic stimulation modulates activation of extinction-related brain regions and enhances contextual extinction learning without affecting renewal. *Front Behav Neurosci* **9**: 34.
Livesey EJ, Don HJ, Uengoer M, Thorwart A. 2019. Transfer of associability and relational structure in human associative learning. *J Exp Psychol Anim Learn Cogn* **45**: 125–142.
Lucke S, Lachnit H, Koenig S, Uengoer M. 2013. The informational value of contexts affects context-dependent learning. *Learn Behav* **41**: 285–297.
Mackintosh NJ. 1975. A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychol Rev* **82**: 276–298.
McClelland JL, Rumelhart DE. 1985. Distributed memory and the representation of

general and specific information. *J Exp Psychol Gen* **114**: 159–197.
Méndez-Couz M, Becker JM, Manahan-Vaughan D. 2019. Functional Compartmentalization of the Contribution of Hippocampal Subfields to Context-Dependent Extinction Learning. *Front Behav Neurosci* **13**: 256.
Milad MR, Wright CI, Orr SP, Pitman RK, Quirk GJ, Rauch SL. 2007. Recall of Fear Extinction in Humans Activates the Ventromedial Prefrontal Cortex and Hippocampus in Concert. *Biol Psychiatry* **62**: 446–454.
Morgan CL. 1903. Other Minds than Ours. In *An Introduction to Comparative Psychology*, pp. 36–59, Walter Scott Publishing, London.
Nevin JA. 2012. Resistance to extinction and behavioral momentum. *Behav Processes* **90**: 89–97.
Nieto J, Mason TA, Bernal-Gamboa R, Uengoer M. 2020. The impacts of acquisition and extinction cues on ABC renewal of voluntary behaviors. *Learn Mem* **27**: 114–118.
Nieto J, Uengoer M, Bernal-Gamboa R. 2017. A reminder of extinction reduces relapse in an animal model of voluntary behavior. *Learn Mem* **24**: 76–80.
Packheiser J, Donoso JR, Cheng S, Güntürkün O, Pusch R. 2020a. Trial-by-trial dynamics of reward prediction error associated signals during extinction learning and renewal. *Prog Neurobiol* 101901.
Packheiser J, Güntürkün O, Pusch R. 2019. Renewal of extinguished behavior in pigeons (Columba livia) does not require memory consolidation of acquisition or extinction in a free-operant appetitive conditioning paradigm. *Behav Brain Res* **370**: 111947.
Packheiser J, Pusch R, Stein CC, Güntürkün O, Lachnit H, Uengoer M. 2020b. How competitive is cue competition? *Q J Exp Psychol* **73**: 104–114.
Pearce JM, Hall G. 1980. A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol Rev* **87**: 532–552.
Pearce JM, Hall G. 1979. The influence of context-reinforcer associations on instrumental performance. *Anim Learn Behav* **7**: 504–508.
Podlesnik CA, Kelley ME, Jimenez-Gomez C, Bouton ME. 2017. Renewed behavior produced by context change and its implications for treatment maintenance: A review. *J Appl Behav Anal* **50**: 675–697.
Podlesnik CA, Sanabria F. 2011. Repeated extinction and reversal learning of an approach response supports an arousal-mediated learning model. *Behav Processes* **87**: 125–134.
Podlesnik CA, Shahan TA. 2009. Behavioral momentum and relapse of extinguished operant responding. *Learn Behav* **37**: 357–364.
Podlesnik CA, Shahan TA. 2010. Extinction, relapse, and behavioral momentum. *Behav Processes* **84**: 400–411.
Redish AD. 2004. Addiction as a Computational Process Gone Awry. *Science* **306**: 1944–1947.
Redish AD, Jensen S, Johnson A, Kurth-Nelson Z. 2007. Reconciling reinforcement learning models with behavioral extinction and renewal: Implications for addiction, relapse, and problem gambling. *Psychol Rev* **114**: 784–805.
Rescorla RA. 1993. Inhibitory associations between S and R in extinction. *Anim Learn Behav* **21**: 327–336.
Rescorla RA. 1997. Response Inhibition in Extinction. *Q J Exp Psychol Sect B* **50**: 238–252.
Rescorla RA, Wagner AR. 1972. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In *Classical Conditioning II:*

Current Research and Theory (eds. A.H. Black and W.F. Prokasy), pp. 64–99, Appleton-Century-Crofts, New York.
Rose J, Otto T, Dittrich L. 2008. The Biopsychology-Toolbox: A free, open-source Matlab-toolbox for the control of behavioral experiments. *J Neurosci Methods* **175**: 104–107.
Schmajuk NA, Moore JW. 1985. Real-time attentional models for classical conditioning and the hippocampus. *Physiol Psychol* **13**: 278–290.
Skinner BF. 1938. *The behavior of organisms: an experimental analysis*. Appleton-Century, Oxford, England.
Starosta S, Stüttgen MC, Güntürkün O. 2014. Recording single neurons' action potentials from freely moving pigeons across three stages of learning. *J Vis Exp* **88**: 51283.
Stollhoff N, Menzel R, Eisenhardt D. 2005. Spontaneous Recovery from Extinction Depends on the Reconsolidation of the Acquisition Memory in an Appetitive Learning Paradigm in the Honeybee (Apis mellifera). *J Neurosci* **25**: 4485–4492.
Tinsley MR, Timberlake W, Sitomer M, Widman DR. 2002. Conditioned inhibitory effects of discriminated Pavlovian training with food in rats depend on interactions of search modes, related repertoires, and response measures. *Anim Learn Behav* **30**: 217–227.
Todd TP. 2013. Mechanisms of renewal after the extinction of instrumental behavior. *J Exp Psychol Anim Behav Process* **39**: 193–207.
Todd TP, Vurbic D, Bouton ME. 2014. Mechanisms of renewal after the extinction of discriminated operant behavior. *J Exp Psychol Anim Learn Cogn* **40**: 355–368.
Trask S, Bouton ME. 2014. Contextual control of operant behavior: evidence for hierarchical associations in instrumental learning. *Learn Behav* **42**: 281–288.
Trask S, Thrailkill EA, Bouton ME. 2017. Occasion setting, inhibition, and the contextual control of extinction in Pavlovian and instrumental (operant) learning. *Behav Processes* **137**: 64–72.
Tulving E, Thomson DM. 1973. Encoding specificity and retrieval processes in episodic memory. *Psychol Rev* **80**: 352–373.
Uengoer M, Lachnit H, Lotz A, Koenig S, Pearce JM. 2013. Contextual control of attentional allocation in human discrimination learning. *J Exp Psychol Anim Behav Process* **39**: 56–66.
Üngör M, Lachnit H. 2006. Contextual control in discrimination reversal learning. *J Exp Psychol Anim Behav Process* **32**: 441–453.
Winterbauer NE, Bouton ME. 2010. Mechanisms of resurgence of an extinguished instrumental behavior. *J Exp Psychol Anim Behav Process* **36**: 343–353.
Wirth S, Yanike M, Frank LM, Smith AC, Brown EN, Suzuki WA. 2003. Single neurons in the monkey hippocampus and learning of new associations. *Science* **300**: 1578–81.
Zentall TR, Wasserman EA, Urcuioli PJ. 2014. Associative concept learning in animals. *J Exp Anal Behav* **101**: 130–151.

Figures and Tables

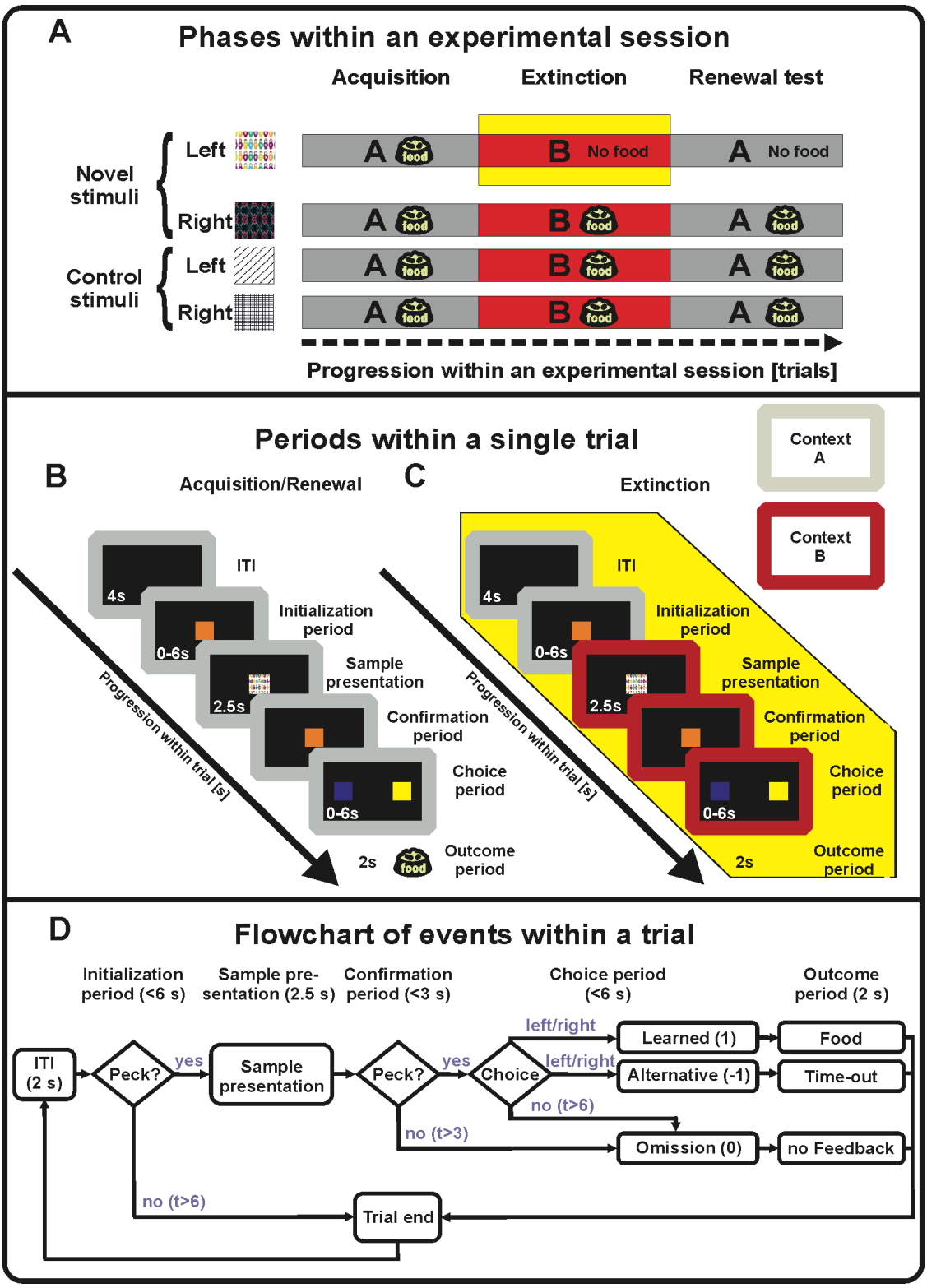

Figure 1. *Experimental design.* (A) Overview of the behavioral paradigm. During the *acquisition phase* (context A) animals learned to associate left/right choices with two novel stimuli. After reaching a learning criterion, extinction started in context B. One novel stimulus was randomly assigned to be extinguished (highlighted in yellow). In the *renewal test phase* context A was re-established, while pecking the extinction stimulus still did not yield reward. (B) Trial structure and sequence of screen presentations during acquisition/renewal test. The transitions between the trial periods are described in panel D. (C) Trial structure and sequence of screen presentations during extinction trials. Initialization triggered a change in the light conditions to context B (red frame), which remained until the end of the trial. In extinction stimulus trials the outcome period remained void of any feedback regardless of the animal's decision (highlighted in yellow). (D) Flowchart depicting transitions between trial periods and possible outcomes. Note that in trials involving the extinction stimulus trials during the *extinction phase* and the *renewal- test phase,* the outcome period remained void of any feedback regardless of the animal's decision (cf. panel C).

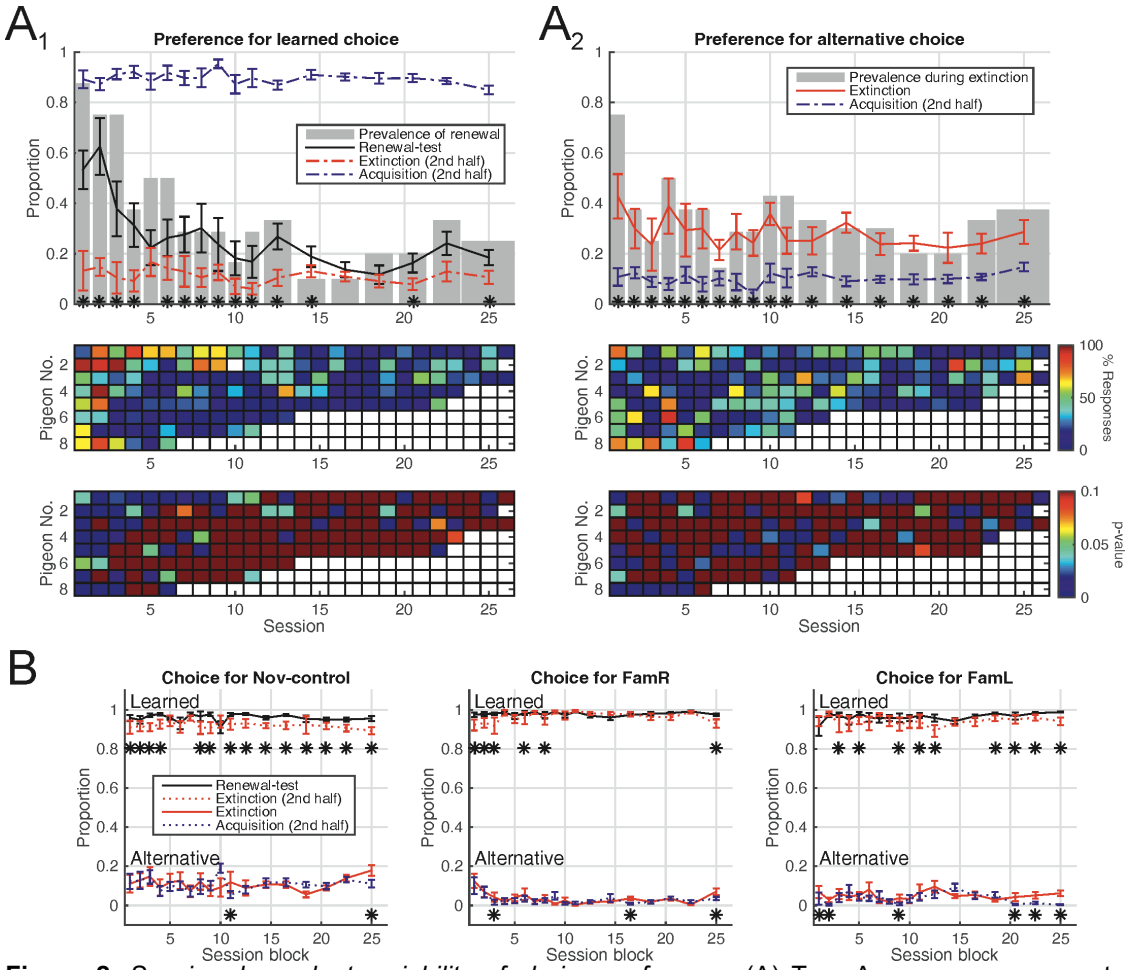

Figure 2. *Session-dependent variability of choice preference.* (A) Top: Average response rate across pigeons (lines) and proportion of pigeons expressing a significant number of responses (bars) as a function of session block. Data from sessions >11 were grouped in blocks of 2 to 3 sessions. Bar widths indicate the size of each block. Stars denote a significant difference ($p<0.05$; z-test) between adjacent traces for each block. Middle: Response rates (color coded) for the learned choice during renewal-test (A1) and for the alternative choice during extinction (A2) for all pigeons and sessions. Bottom: p-values (color coded) of the respective response rates shown in middle for all pigeons and sessions. During renewal-test, both the preference and prevalence of

the learned choice (black trace and gray bars in A1, respectively) decay with session number. During extinction (A2), the proportion of pigeons exhibiting a preference for the alternative choice is significantly higher during the first session. (B) Average response rate across pigeons for the novel control stimulus (left), and two familiar stimuli (middle and right).

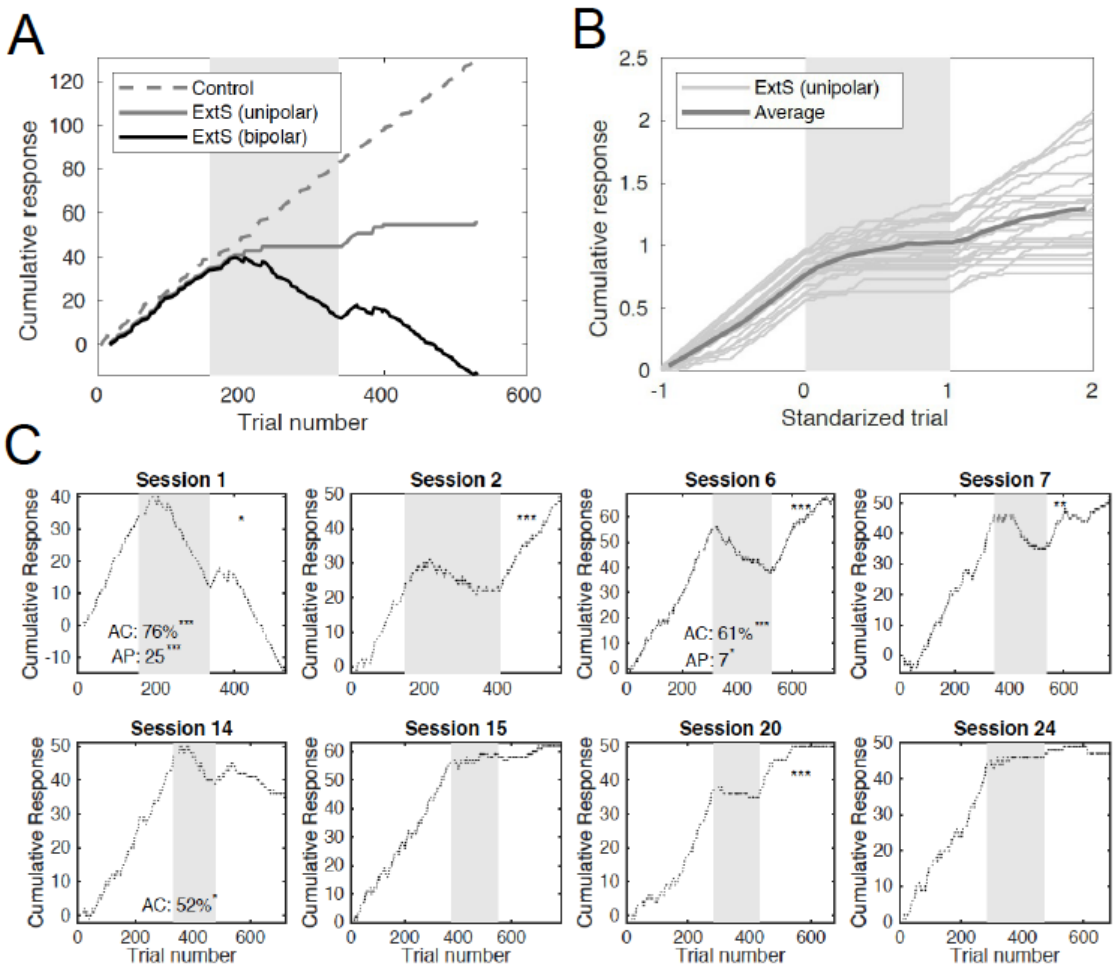

Figure 3. *Visualization of choice behavior in single sessions.* (A) Single session learning curve obtained using two different choice codings. The trial choices (learned, omission, alternative), can be encoded in a unipolar (1, 0, 0) or bipolar (-1,0,+1) fashion. Unipolar coding (solid gray trace) shows the decay of the learned response in the extinction phase (gray area) and the renewal of the learned response upon return to context A. The bipolar coding (black trace) uncovers the choice behavior during the extinction phase, where a negative slope shows preference for the alternative choice over omissions in the extinction and renewal-test phase. Responses to control stimuli (dashed gray trace) in interleaved trials remained consistent throughout the session. (B) Cumulative record of unipolar-encoded curves for all sessions (thin traces) obtained from one animal along with the grand average (thick trace). Trial numbers are standardized for visualization and averaging (Acquisition: [-1 0); Extinction: [0 1); Renewal: [1 2)). (C) Bipolar-encoded cumulative learning curves for a sample of eight sessions from one animal. Proportion of alternative-choice (AC) responses and alternative persistence (AP; number of consecutive trials yielding an alternative choice) during extinction (gray area) are shown when significant ($p < 0.05$). Stars on the top-right corner of panels mark the significance of the renewal effect.

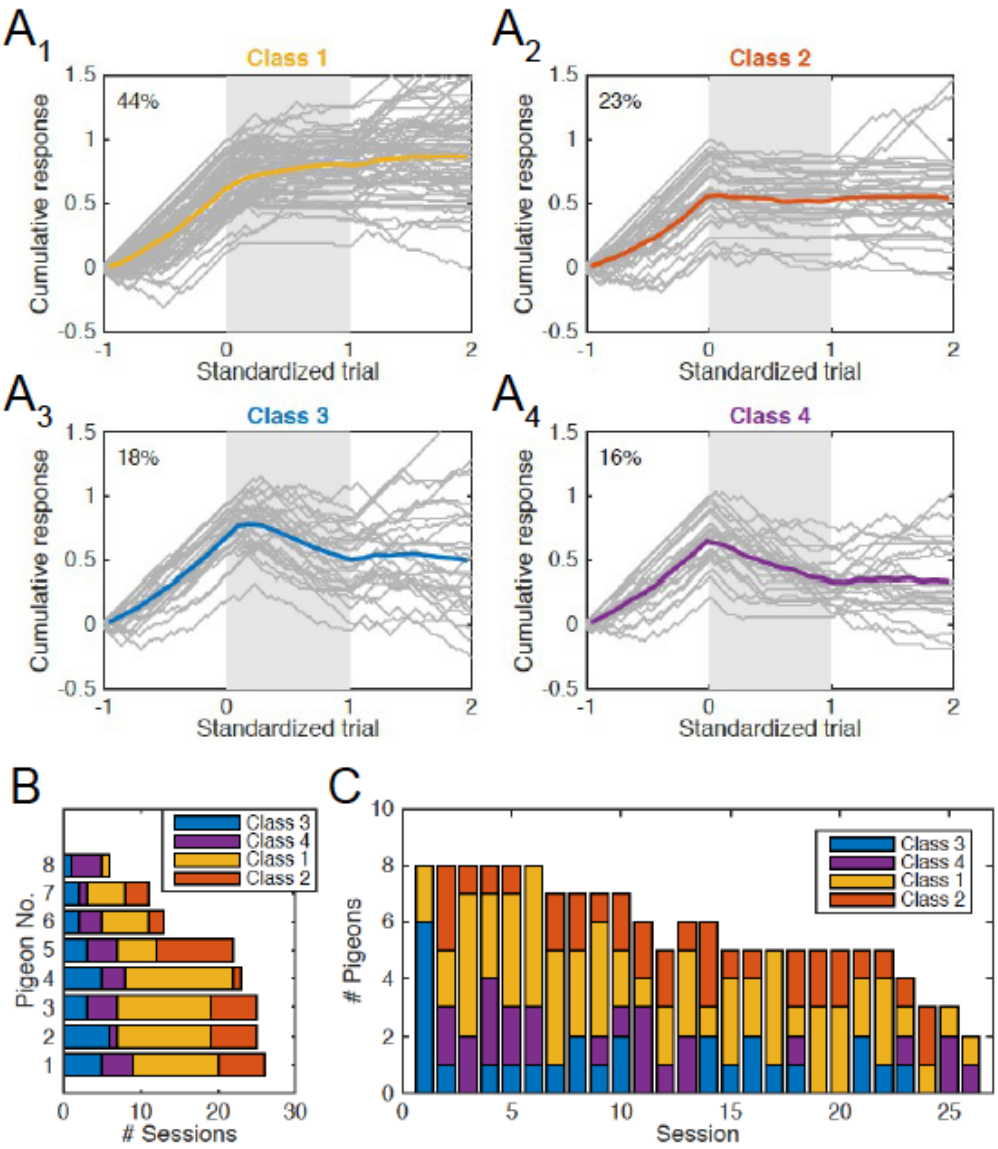

Figure 4. *Variability of behavior during extinction.* (A) Standardized learning curves (gray traces) corresponding to 151 behavioral sessions obtained from 8 animals. For all sessions, -1 represents the onset of acquisition, 0 the onset of extinction, 1 the onset of the renewal test and 2 the end of the experiment. Curves were classified according to their mode of transition from the acquisition to the extinction phase (smooth vs. abrupt) and their expression of alternative choices during the extinction phase. A1: smooth transition and no alternative choice; A2: abrupt transition and no alternative choice. A3: smooth transition and alternative choice. A4: abrupt transition and alternative choice. Number at the top left corner of each panel indicates the proportion of learning curves that fall into the respective class. (B) Number of learning curves (# sessions) that fall into each of the four classes for each animal. Individual pigeons do not exhibit a clear bias for a particular type of learning curve. (B) Number of animals expressing each class of learning curve across sessions. During the first session, all pigeons exhibited smooth transitions at the onset of context B (Classes 1 or 3). Abrupt transitions (Classes 2 or 4) emerged exclusively after the first session. (C) Number of sessions in which significant renewal is observed, for each animal.

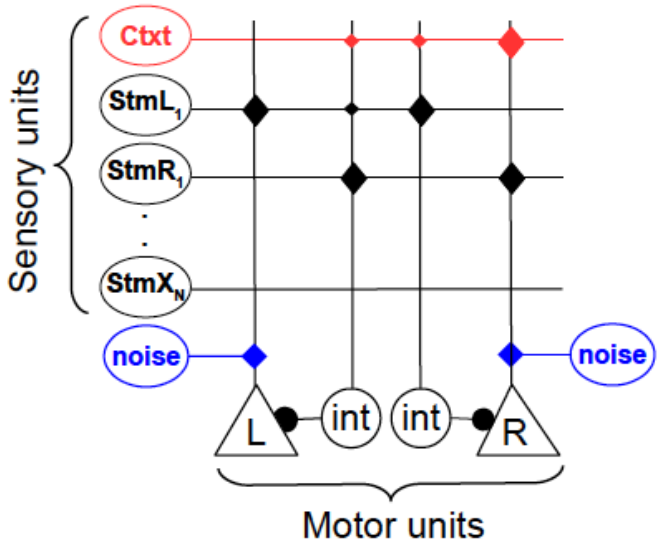

Figure 5. Associative network model. Sensory units (ovals) can establish excitatory associations directly with motor units (triangles) mediating the left and right responses, or inhibitory associations via interneurons (circles). Context is treated as another simultaneously presented stimulus (red oval). Motor units also receive excitatory noise (blue ovals). Synapses (diamonds) mediating excitatory and inhibitory associations are reinforced every time a reward is delivered, or not, respectively. The motor unit receiving the strongest net input generates the respective response. If both units are inhibited below threshold, an omission ensues.

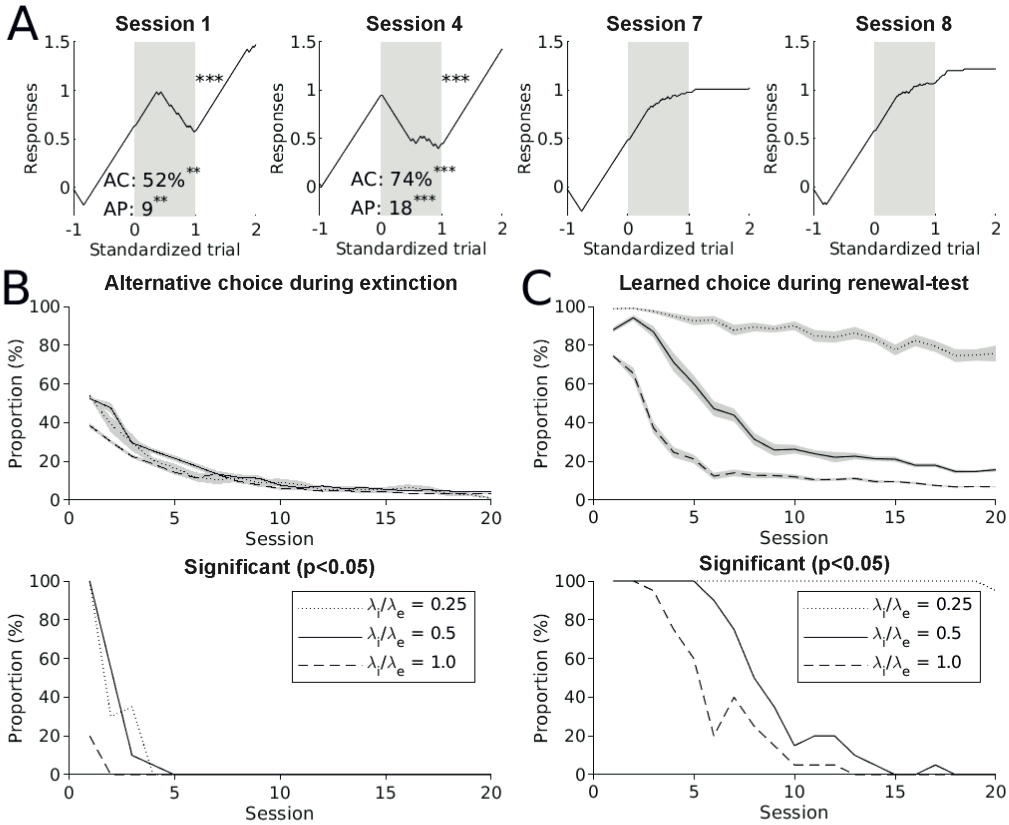

Figure 6. *Associative learning accounts for the features observed in the behavior of pigeons in extinction learning and renewal.* (A) Sample sessions obtained from one pigeon-model, showing strong preference for the alternative choice (sessions 1 and 4), an abrupt transition at the onset of the extinction phase (session 4) and decay of the renewal effect (sessions 7 and 8). (B) Preference for the alternative choice during extinction as proportion of emitted choices (top) and proportion of pigeon-models emitting a significant number of alternative choices (bottom). (C) Prevalence of the renewal effect expressed as proportion of emitted learned choices (top) and proportion of pigeon-models emitting a significant number of learned choices (bottom) during the renewal-test phase. Model results were obtained from a batch of 20 pigeon-models subjected to 20 sessions with randomly selected contexts and extinction stimuli. Simulations were run using three different learning rates for inhibitory connections (0.005, 0.01 and 0.02).

Supplementary Information

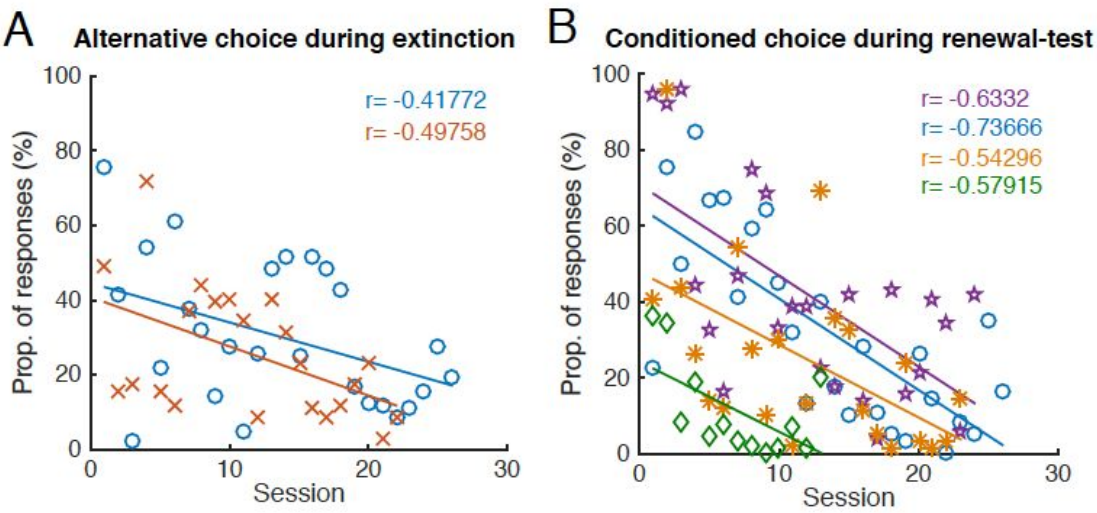

Figure S1. *Preference for target response as a function of session number.* (A) Preference for alternative choice during extinction. (B) Preference for conditioned choice during renewal. Only pigeons showing significant correlations are shown (symbols). Correlations for the corresponding linear fits (colored lines) are indicated in the top corner of the panel (colored text).

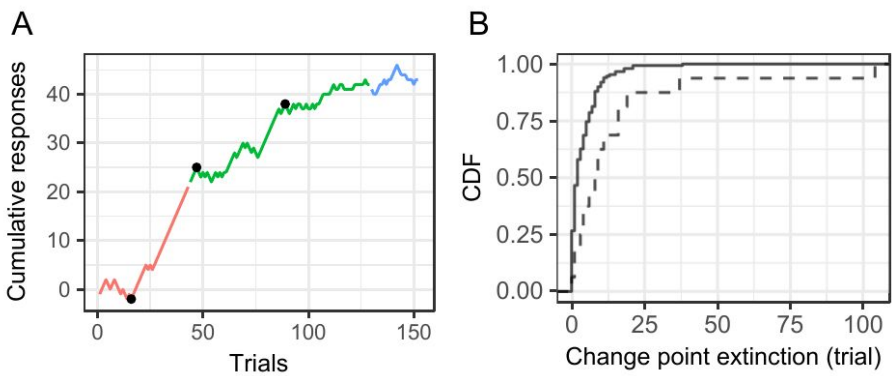

Figure S2. *Change point analysis reveals changes of behavior across sessions*. (A) Cumulative responses to the extinction stimulus in a single session. Phase is color-coded (red: acquisition; green: extinction; blue: renewal-test). Black dots show the change-points as identified by the change point analysis. (B) Cumulative distribution of the trial number, at which the change point occurs during the extinction phase (dashed: first session; solid: remaining sessions). The change point occurs earlier in later sessions.

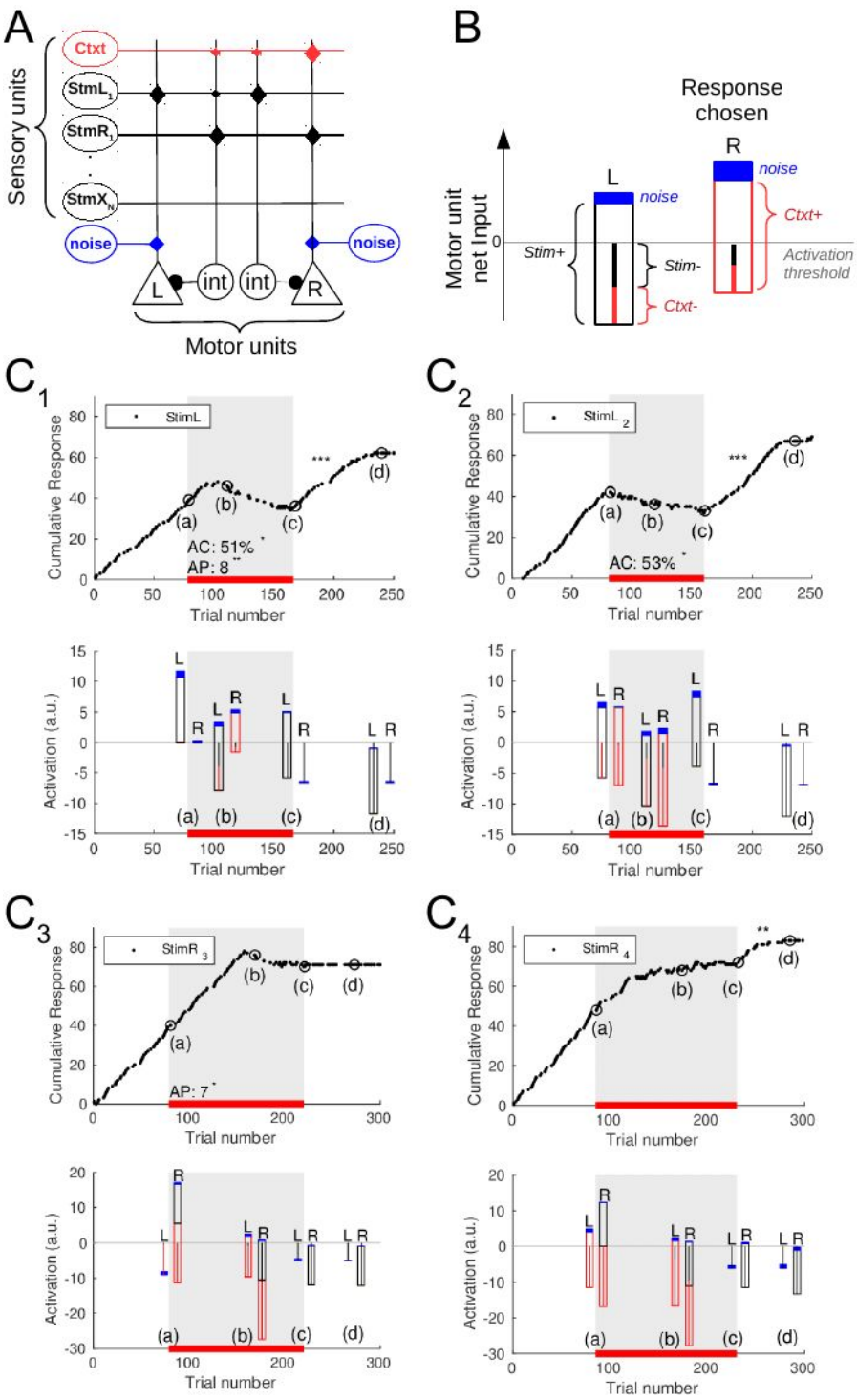

Figure S3. *Associative learning can drive complex choice-behavior during extinction and renewal.* (A) Associative network. Sensory units (ovals) can establish excitatory associations directly with motor units (triangles) mediating the left and right responses, or inhibitory associations via interneurons (circles). Motor units also receive excitatory noise. (B) Schematic of how the composition of the motor unit input activity is depicted: Inhibition (indicated by vertical

lines), excitation (indicated by open bars), and excitatory noise (indicated by solid blue bars). Hence, the net activation is indicated by the top of the bar. The unit with the highest net activation triggers the corresponding behavioral response. If the net activation of both motor units remains below the threshold, a choice omission ensues. (C) Model responses to four consecutive sessions of a simplified version of the task performed by the pigeons. (top) Cumulative responses for the extinction stimulus. The proportion of alternative choices (AC) and longest chain of successive alternative choices (AP: Alternative persistence) during extinction are shown. (C1) Preference for the alternative choice during extinction. (C2) Abrupt transition upon onset of extinction context. (C3) Absence of renewal. (C4) Reappearance of renewal. Note that the variable dynamics of extinction emerged due to remnant context-response associations from previous sessions. (bottom) Input composition of the left and right motor-output units (L and R, respectively) in response to the extinction stimulus (black) and context (red). The contribution of each component to the activation is coded as shown in B. Activity is sampled at the onset of the extinction phase (a), during extinction (b), at the onset of the renewal-test phase (c), and at the end of the renewal-test phase (d); see corresponding circles on top. The learning rates of all connections were set to 0.02, and all synaptic weights saturated at a value of 20 (see Materials and Methods).

Supplementary Text

How associative networks can drive complex behavior. To illustrate the basic interactions underlying the behavior of the model, we put the model through a simplified version of the protocol used in the behavioral experiments (Fig. S3). Here, only two stimuli were presented instead of four, and only one extinction context was used instead of two. Briefly, during the acquisition phase, responses to the left or right were reinforced when the corresponding stimuli, StimL or StimR, were presented, respectively. At the onset of the extinction phase, the context unit was activated and no rewards were given in the presence of the extinction stimulus (StimL in Fig. S3C1), regardless of the response given. Figure S3C shows the behavior of the model during the first and subsequent three sessions of the task. These four sessions provide an example of how associative learning can generate the complex choice behavior observed in the experimental data. To illustrate the evolution of the associations giving rise to the behavior of the model, we display the excitatory and inhibitory contributions of both context and extinction stimulus to the activity of the left and right motor units (Fig. S3B and Fig. S3C, bottom panels) at specific points of the learning curve (indicated by red circles in Fig. S3C, top panels). Namely, at the onset of the extinction phase (a), during extinction (b), at the onset of the renewal-test phase (c), and at the end of the renewal-test phase (d).

<u>No abrupt transition at context B onset in first sessions.</u> In the first session, the resulting cumulative response to the extinction stimulus, e.g. StimL, exhibits the expected positive slope at the end of acquisition (Fig. S3C1, black trace), which remains for several trials during the extinction phase (indicated by the gray shaded area and the horizontal red bar in Fig. S3C). This smooth transition at the onset of context B (Fig. S3C1, (a)) is driven by the strong association between StimL and the left response that was established during the acquisition phase, which is reflected in the strong net activation of the left motor unit due to the presence of StimL (Fig. S3C1 (a) in the bottom panel). As the extinction phase progresses, the conditioned choice is gradually suppressed, since the lack of reinforcements to operant responses to StimL leads to the emergence of negative associations between StimL and the left response, and between StimL and context (Fig. S3C1 (b) black and red lines in bottom panel, respectively). Since it takes several trials without reinforcements to build up the inhibition required to suppress the activation of the L response in the presence of StimL, our model predicts that the choice behavior changes smoothly at the onset of the extinction phase in the first session.

<u>Preference for the alternative choice during extinction.</u> As extinction learning progresses, the

model favors the alternative choice over omissions (Fig. S3C1, (b)), just as pigeons often did (see Fig. 1B, 2A). This behavior results from the higher activation of motor unit R in the presence of the extinction stimulus StimL, which is explained as follows: In our experimental paradigm, during the extinction phase, responses of the motor unit R have been reinforced in the presence of both StimR (non-extinction stimulus) and the extinction context (red illumination). Consequently, a positive association between context and the right motor unit (Fig. S3C1 (b), red box) has formed. This association between the context and the motor unit R alone is now strong enough to tilt the balance between the two responses in favor of the alternative (right) choice in the presence of the extinction stimulus StimL. As the extinction phase progresses, responses in the presence of StimL remain unrewarded. Therefore, all the inhibitory connections from StimL and context to both, the left and right motor units, are further reinforced. This example illustrates the principle behind how alternative choices arise in our model. First, there is competition between excitatory and inhibitory drive to execute a particular response. Second, the different response options (R or L) compete against one another. The model's choice is ultimately the outcome of these two levels of competition in the decision-making process.

Abrupt changes of behavior at the onset of the extinction phase. In the second session, during acquisition, a new pair of stimuli, $StimL_2$ and $StimR_2$, are associated with the left and right responses, respectively, and $StimL_2$ was chosen as the extinction stimulus. This time, upon the onset of the extinction phase, there is a positive association between context and the motor unit R, which was established in the previous session (Fig. S3C2 (a), bottom; compare with the corresponding point in Fig. S3C1). As a result, in this example, the net activation of L and R is nearly balanced, and the given response is mostly determined by the noise. In the general case, however, it is also possible that the activation of R stemming from the positive influence of context is able to tilt the balance in favor of the alternative choice (the right response in this example). In any case, an abrupt change in behavior upon the onset of context B ensues (Fig. S3C2 (a)). Since this mechanism requires previous exposure to the extinction context, this behavior could not be observed during the first session in the model, similar to our finding in pigeons (Fig. 3C).

Intermittence of the renewal effect. In sessions 1 and 2, it is possible to see how operant responses to the extinction stimulus suddenly re-emerge upon the onset of the renewal-test phase (Fig. S3C1 and S3C2 (c)). Here, renewal emerges due to the release of inhibition by context B on a specific response (Fig. S3C1 and S3C2, bottom (c)), as previously suggested (Todd 2013; Todd et al. 2014). However, this effect vanishes in the third session (Fig. S3C3, top (c), and reappears in the fourth (Fig. S3C4, top, (c)). This intermittence of the renewal-effect is explained as follows: In the third session (Fig. S3C3), $StimR_3$ is chosen as the extinction stimulus. At the onset of extinction, there is a very strong activation of the right response (Fig. S3C3, bottom (b)) due to its positive association to both $StimR_3$ and context B, established during acquisition and previous sessions, respectively. Therefore, it takes more trials to extinguish the association between $StimR_3$ and the right response, which translates to a persistent extinction curve (Fig. S3C3, top, gray area). This long process of extinction, in turn, results in a very strong negative association between $StimR_3$ and the right response at the end of the extinction phase (Fig. S3C3, bottom (b)). Therefore, at the onset of the renewal-test, the release of context inhibition on the right response is not sufficient to drive the renewal effect (Fig. S3C3, bottom (c). In the fourth session, however, the strength of both negative and positive associations between context and the right response are close to their balanced saturated state (Fig. S3C4, bottom, (a)). Thus, the extinction process increases only the negative association between $StimR_4$ and the right response, leaving the remaining context-right response associations intact (Fig. S3C4, bottom, (b)). In this case, at the onset of the renewal-test phase, the net input to the right response resulting from the negative and positive associations between $StimR_4$ and the right response established during acquisition and extinction is still positive, resulting in the

re-emergence of renewal (Fig. S3C4, (c)).

In the example sequence shown in Figure S3C, the associations between the extinction context and the motor response that is no longer rewarded during extinction created imbalances at the initial stage of extinction (Fig. S3C, points a). These imbalances gave rise to counterintuitive behaviors in sessions 1 to 3. However, once both positive and negative associations between the context and L and R responses balance each other out, the context can no longer exert its counter-intuitive effect on the responses. Such wearing-out of the context effectiveness not only predicts a more prominent preference for alternative choices during the first session, but also an overall decay of the renewal effect.

Different rates of extinction learning. Extinction in our model occurs at different speeds in different sessions (Fig. S3), even though the learning rate parameter in our model is held fixed across sessions. The varying extinction speed is an emerging phenomenon that is driven by two factors. First, the learning rule adjusts weights proportionally to the difference between the current weight and the saturation weight. Since weights are maintained across sessions, they start at different levels in different sessions. As a result, the weight adjustments, and in turn the learning speeds, differ. Second, the emitted response is determined in a competition between the influences of the stimulus and that of the context. Combined with the first point, the associative strength depends on the history of previous and concurrent exposures. For instance, if the context had not previously acquired much associative strength, the conditioned response can be suppressed quite quickly by adding negative associative strength to the context (Fig. S3 C1). If on the other hand, the context's overall associative strength was positive and its magnitude higher than that of the overall negative associative strength of the extinction stimulus initially, the model continues to emit conditioned response for a longer period during extinction until they are suppressed by negative associations with the context (Fig. S3 C3).