Complex dynamics of choice during operant extinction and a "simple" associative learning account

José R. Donoso[1], Julian Packheiser[2], Roland Pusch[2], Zhiyin Lederer[1], Thomas Walther[1], 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.

* Correspondence to: sen.cheng@rub.de

Abstract

Extinction learning, the process of ceasing an acquired behavior in response to altered reinforcement contingencies, is essential for survival in a changing environment. So far, research has mostly neglected the learning dynamics and variability of behavior during extinction learning and instead focused on a few response types that were studied by population averages. Here, we take a different approach by analyzing the trial-by-trial dynamics of operant extinction learning in both pigeons and a computational model. The task involved discriminant operant conditioning in context A, extinction in context B, and a return to context A to test the context-dependent return of the conditioned response (ABA renewal). By studying single learning curves across animals under repeated sessions of this paradigm, we uncovered a rich variability of behavior during extinction learning: (1) Pigeons prefer the unrewarded alternative choice in one-third of the sessions, predominantly during the very first extinction session an animal encountered. (2) In later sessions, abrupt transitions of behavior at the onset of context B emerge, and (3) the renewal effect decays as sessions progress. While these results could be interpreted in terms of rule learning mechanisms, we show that they 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 unexpected power of "simple" associative learning processes.

Significance Statement

Operant conditioning is essential for the discovery of purposeful actions, but once a stimulus-response association is acquired, the ability to extinguish it in response to altered reward contingencies is equally important. These processes also play a fundamental role in the development and treatment of pathological behaviors such as drug addiction, overeating and gambling. Here we show that extinction learning is not limited to the cessation of a previously reinforced response, but also drives the emergence of complex and variable choices that change from learning session to learning session. At first sight, these behavioral changes appear to reflect abstract rule learning, but we show in a computational model that they can emerge from "simple" associative learning.

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

Author Contributions: Designed research: JRD, JP, RP, 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, OG, SC.

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 (1). 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 (2–6). Once a purposeful behavior is acquired, the ability to extinguish it as a result of altered reward contingencies is also essential for survival. The importance of this so-called extinction learning is emphasized by the fact that all vertebrate and invertebrate species tested exhibit this ability (7–14).

Although extinction may involve some erasure of the previously acquired memory (15, 16), there is strong evidence that it also involves new learning (17–19). According to the latter view, during extinction, the previously acquired memory trace 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 (20). A prominent phenomenon in support of extinction as new learning is the renewal effect, where an extinguished behavior reemerges when the animal is removed from the context where extinction learning took place (17, 18, 21). If extinction were constituted by erasure of the acquired association, a context-dependent relapse of the previously conditioned response should not occur. Maintaining the acquired association despite extinction might offer an advantage over deletion since behavior that was once useful might become useful again in the future under different conditions. However, the 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 (22). Therefore, understanding operant extinction and the factors that influence its reappearance can be helpful in developing treatments for several pathological behaviors and in preventing their reappearance (2, 22).

Previous studies have shown that extinction is not limited to a decrease in a previously reinforced behavior, but it can also drive the emergence of new, previously non-reinforced behaviors (23–27). This becomes particularly relevant in real world settings and experimental designs where there are multiple alternative choices that have equivalent value in a particular situation. Notwithstanding, studies of operant extinction have focused mostly on the cessation of specific responses, often neglecting the effect that extinction might have on other available actions. Consider, for example, a rat in a T-maze that has been conditioned to choose the right arm in response to a discriminant stimulus. During a subsequent extinction phase, this rat can choose between at least four equally non-reinforced behaviors: turn left, turn right, return to the starting point before reaching the decision point, and refuse to leave the starting point altogether. Nevertheless, a typical analysis focuses only on the cessation of the previously conditioned response, in this case, turning right (28, 29). In this way, the effect of extinction learning on choice behavior remains undisclosed. Taking into account this choice behavior and its variability is important to test putative mechanisms of extinction learning and decision making. One particularly challenging issue in this respect is to explain the expression of alternative choices that provide no added value, such as in the example case described above. In fact, existing models 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 (15, 30).

A further major problem for understanding the cognitive processes and neural mechanisms underlying operant extinction is the tradition of pooling and averaging across subjects and sessions. Such analyses not only conceal the complexity of behavior during a learning task (28,

31–34), but also obscure the dynamics of learning within and across sessions (18, 19). Considering this variability is of particular importance for three main reasons. First, some important features of the behavior, such as the expression of alternative choices during extinction, or the absence of the renewal effect in some sessions or subjects (31) can be lost in the grand average. Second, computational models usually assume that the properties of the average curve are representative of those of individual curves (30, 32, 33). 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. Third, relevant neural correlates of behavior can be missed or misinterpreted if neural activity and behavior are averaged across trials or sessions, or if neural activity is averaged across neurons, since the averages are probably not representative of single-trial outcomes (34, 35).

In the present study, we analyzed the choice behavior of pigeons undergoing multiple sessions of an operant extinction learning task. Each session consisted of three subsequent stages: (1) acquisition learning of discriminant operant behavior in context A; (2) extinction learning in context B, and (3) return to context A to test for ABA renewal. Furthermore, we addressed the aforementioned issues of classical behavioral analyses by focusing 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. Considering these diverse and individual learning trajectories, we developed a parsimonious model consisting of an associative network and a winner-takes-all decision making process. In spite of its simplicity, this model could account for the rich behavioral phenomena observed in the data.

Results

Animals underwent multiple sessions of a discriminative operant conditioning task that consisted of three consecutive phases (Fig. 1A) (36). In the initial acquisition phase, which took place under white house lights (context A), animals had to learn to associate two session-unique novel visual stimuli with either a left-peck or a right-peck response. During acquisition, food was delivered after every correct choice. Once animals reached a performance threshold for both stimuli, the task entered the extinction phase, which took place under colored house lights, either red (context B1) or green (context B2). Here, one of the novel stimuli was randomly chosen to become the extinction stimulus (ExtS). Responses to this stimulus were no longer rewarded for the remainder of the session. Once the responses to the extinction stimulus dropped below a performance threshold, the task switched to the renewal-test phase with a return to context A. This third phase was used to test whether operant responses associated with the extinction stimulus reappeared under context A in the absence of reinforcement. Throughout the whole session, trials in which the novel stimuli were presented were interspersed with control trials, in which two other familiar stimuli were presented and in which pigeons were consistently rewarded for correct responses. These two stimuli served both as controls as well as fix points for the animals throughout the experiment as they did not have to learn their stimulus-response association in each individual session. In total, we collected data for 156 sessions obtained from 12 pigeons, and we analyzed their behavior as follows.

Visualizing choice-behavior in single sessions. To visualize the time-course of learning within single sessions, we focused on the cumulative record of successive behavioral responses as a function of trial number (37, 38). However, we departed from the traditional ‘unipolar’ way of encoding the responses, which focuses on the presence of the reinforced choice only. Instead, we used a ‘bipolar’ encoding to reveal the presence of alternative choices (Fig. 1A, middle). Figure 1B illustrates the difference between the cumulative learning curves resulting from these

two ways of encoding the behavior. In the traditional unipolar encoding, the presence or absence of the reinforced response in a given trial is signaled by 1 or 0, respectively. Thus, the expression of the conditioned response is reflected in the cumulative learning curve as a positive slope, and extinction is revealed as a gradual decay of the slope towards 0. This pattern is visible in individual sessions (Fig. 1B, gray trace) as well as in the grand average across sessions (Fig. 1C). The presence of the renewal effect, on the other hand, is marked by a sudden increase in the slope of the cumulative learning curve upon switching to the acquisition context A. The grand average shown in Figure 1C reflects this received view on extinction learning and is consistent with previous studies on ABA renewal (2, 9, 31, 36). However, the unipolar encoding focuses only on the response that has been conditioned during the acquisition phase (henceforth, 'conditioned choice'), thus occluding the effect of extinction on the other available response (henceforth, 'alternative choice'). To gain a more detailed view of the behavior within a session, we encoded each conditioned choice, alternative choice, or omission as +1, -1 and 0, respectively. Using this bipolar encoding, the slope in the cumulative curve reveals biases towards specific responses (Figs. 1 B and 2A, black traces): a positive or negative slope indicates a tendency to prefer the conditioned choice or the alternative choice, respectively. A slope of 0 indicates either a continuous chain of omissions or random mixtures of conditioned and alternative choices, which could be interpreted as exploratory responses. In Figure 1B, for example, the bipolar encoding (black trace) reveals that extinction and re-extinction are dominated by a persistent selection of the alternative choice, even though the animal was never rewarded for this response in the presence of the extinction stimulus. This behavior was hidden in the cumulative record when both, alternative choices and omissions, were encoded as 0 in the unipolar encoding (Fig. 1B, gray trace). Thus, the use of the bipolar encoding uncovered a rich variability of choice behavior during the extinction and renewal-test phases (Fig. 2).

Diversity and dynamics of choice-behavior during extinction. According to the canonical 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 (39–41). In our analysis, learning curves were considered canonical (e.g., Fig 2A1) if behavior remained unchanged within the first 5 presentations of the extinction stimulus, and if they exhibited no significant preference for the alternative choice during the extinction phase (see Materials and Methods). We grouped all the sessions exhibiting these features under class 1 (Fig. 2B1), which accounted for 44% of the cumulative learning curves across all pigeons and sessions. However, in other cases, even though extinction was dominated by omissions as in the canonical case, the change in behavior seemed to occur abruptly upon the onset of context B (e.g., Fig. 2A2). These changes were apparently driven by the switch from the acquisition to the extinction context. To quantify the presence of these abrupt transitions, we considered a transition abrupt if the 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). All the sessions exhibiting abrupt transitions and no significant preference for the alternative choice (see below) were grouped as class 2, and constituted 22% of the curves analyzed (Fig. 2B2).

Furthermore, we also found curves that developed negative slopes during extinction, revealing a preference for the alternative choice over omissions (e.g. Figs. A3 and A4), even though this behavior was never reinforced for the extinction stimulus. To assess the significance of these responses, we calculated the probability p of obtaining the observed number of alternative choices by chance (see Methods for details), and regarded the observation as significant if $p<0.05$. According to this analysis, in one-third of the sessions animals exhibited a significant preference for the alternative choice during the extinction phase. Among these sessions, 90% exhibited chains of 5 to 25 consecutive trials (median: 7 trials) where the animals opted for the

alternative choice, further indicating that these choices are not random occurrences. We name these chains persistent alternative choices. We further divided the extinction sessions expressing a significant preference for the alternative choice according to whether they expressed smooth transitions upon the onset of context B or not, giving rise to classes 3 (occurrence rate: 18%) and 4 (occurrence rate: 15%), respectively (Figs. B3 and B4). In summary, while 44% of the learning curves followed the canonical view of extinction learning, the majority of the learning curves (56%) deviated in at least one major aspect: either because animals exhibited abrupt transitions at the onset of context B (class 2), or favored the alternative choice over omissions during extinction (class 3), or both (class 4, Fig. 2B4).

The previous analysis focused on the prevalence of diverse behavioral patterns across animals and sessions. Such analysis, however, overlooks not only interindividual variability, but also the effect that learning history might have on the expression of different classes of behavior. To this end, we analyzed the way in which the behavioral types described in Figure 2 were distributed across pigeons and sessions (Fig. 3). To assess interindividual variability, we quantified the prevalence of the four types of learning curve within single pigeons (Fig. 3A), and found that they do not express a particular behavioral type consistently. In fact, all 7 pigeons for which sufficient data was available (> 6 sessions) exhibited all four types of behavior. 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 types (Fig. 3B). Curves with smooth transitions (clusters 1 and 3) dominated the first session (see also Fig. S1), and occurred less frequently in later sessions. To confirm this observation, we used the method proposed by Gallistel et. al. (37) to determine the change point of the learning curves during extinction. We found that the change of behavior occurred significantly later in the first sessions as compared to the subsequent sessions ($p = 0.023$; KS-Test, Fig. S2). Another peculiar aspect of the first session was its relatively large proportion of negative responses (58%), with respect to the proportion found in all the remaining sessions combined (31%; $p=0.028$; z-test).

Variability in the renewal effect. To assess the presence of the renewal effect in a given session, we tested whether the number of conditioned choices emitted during the renewal-test phase was significant or not (Fig. 2A, stars on top right of the panels). Using this analysis, we quantified the prevalence of the renewal effect in each pigeon (Fig. 3C) and its session-to-session variability across pigeons (Fig. 3D). Seven out of 8 pigeons that contributed at least 6 sessions exhibited renewal in a significant ($p < 0.05$; binomial test) fraction of their learning curves (range: 22 % to 67 %). Remarkably, even though the renewal effect appeared intermittently in single pigeons (Fig. S3), its overall prevalence across all pigeons decayed as sessions progressed (Fig. 3D and S3). Indeed, the fraction of pigeons expressing renewal was negatively correlated with session number ($r = -0.68$; $p = 0.002$; Fig. S3).

Associative learning can generate complex choice behavior. 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. In light of these results, it is tempting to conclude that the animals adapt their behavior to the structure of the task when subjected to it repeatedly (see Discussion for details). However, several studies indicate that pigeons might not possess the cognitive capabilities to grasp abstract rules (42–44). Motivated by this apparent inconsistency, we assessed to which extent simple associative learning could account for the behavior we observed. To this end, we implemented a parsimonious model aimed at capturing the associative aspect of the task (Fig. 4A). This model embodies two fundamental principles of associative

learning. First, representations of the present (discriminant) stimuli can freely establish both positive and negative 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 (17, 18, 21). Figure 4A 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 (45, 46). 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 (15, 47, 48). In the following, we use this model to show how the associations between the context signal and the left and right responses established within the time course of one session can give rise to a variety of choice behavior as sessions progress (Fig. 4C). Additionally, we show that this simple associative model, when subjected to the same experimental paradigm as the pigeons in our study, generates a trend of behavioral changes similar to that observed in the pigeons (Fig. 5).

To illustrate the basic interactions underlying the behavior of the model, we begin by putting the model through a simplified version of the protocol used in the behavioral experiments (Fig. 4C). Here, for simplicity, 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. 4C1), regardless of the response given. Figure 4C 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. 4B and Fig. 4C, bottom panels) at specific points of the learning curve (indicated by red circles in Fig. 4C, 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).

No abrupt transition at context B onset in first sessions. 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. 4C1, black trace), which remains for several trials during the extinction phase (indicated by the gray shaded area and the horizontal red bar in Fig. 3B and D). This smooth transition at the onset of context B (Fig. 3C1, (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. 3C1 (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. 3C1 (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.

Preference for the alternative choice during extinction. As extinction learning progresses, the model favors the alternative choice over omissions (Fig. 4C1, (b)), just as pigeons often did (see Fig. 1B, 2A3 and 2A4). 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. 4C (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. 4C2 (a), bottom; compare with the corresponding point in Fig. 4C1). 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. 4C2 (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. 3B).

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. 4C1 and 4C2 (c)). Here, renewal emerges due to the release of inhibition by context B on a specific response (Fig. 4C1 and 4C2, bottom (c)), as previously suggested (18, 21). However, this effect vanishes in the third session (Fig. 4C3, top (c), and reappears in the fourth (Fig. 4C4, top, (c)). This intermittence of the renewal-effect is explained as follows: In the third session (Fig. 4C3), $StimR_3$ is chosen as the extinction stimulus. At the onset of extinction, there is a very strong activation of the right response (Fig. 4C3, 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. 4C3, 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. 4C3, 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. 4C3, 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. 4C4, 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. 4C4, 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. 4C4, (c)).

In the example sequence shown in Figure 4C, 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. 4, 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 (see below).

Associative learning predicts the general trend observed in the behavioral data. So far, we have used a simplified version of the behavioral task to illustrate the principles by which simple associative processes can give rise to some of the complex behaviors observed in the data, namely, we have used only two stimuli in each session, and only one context (B). Furthermore, we have only taken the extinction context into account, omitting the action of context A during the acquisition and renewal-test phases. The inclusion of those elements in the model enables context A to establish positive and negative associations with the response units during the acquisition and renewal-test phases, respectively. These associations are carried over to subsequent sessions, increasing the complexity of the interactions between contexts and responses. In the following, we included these additional features and tested whether a population of pigeon-models subject to several sessions of training could reproduce the trends exhibited by the pigeons in our study (Fig. 2). In particular, we focused on the distribution across sessions of two features of the learning curves, namely, the preference for the alternative choice during the extinction phase, and the vanishing renewal effect.

The population consisted of 20 pigeon-models, each of which was subject 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: The learning rates of all the excitatory connections to motor units (λ_e) were set to 0.02. For the connections to the interneurons mediating the inhibitory associations, the learning rates (λ_i) were set to either to 0.005, 0.01 or 0.02. Thus, we tested 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 (see Methods). Since no parameter was adjusted between sessions, the variability in the learning curves stemmed solely from the history of learning, i.e. the sequence of extinction contexts and extinction stimuli across sessions. A sample of four learning curves obtained from one of the pigeon-models (Fig. 5A) exhibits all three interesting features we uncovered in the pigeon behavior: preference for alternative choices during extinction, abrupt transitions of behavior upon onset of context B, and absence of the renewal effect. 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. 5B, top). The proportion of pigeon-models emitting a significant number of alternative responses (Fig. 5B, bottom) qualitatively reproduces the findings in our experimental data. 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 proportion of conditioned choices emitted in the renewal-test phase (Fig. 5C, top), and in the proportion of pigeon-models emitting a significant number of conditioned choices (Fig. 5C, bottom). The decay

of the renewal effect was strongly modulated by the learning rate of inhibition**.**

Discussion

We have analyzed the behavior of pigeons subject to multiple sessions of a task involving discriminant operant conditioning in context A, extinction in context B, and a return to context A to test for the return of the conditioned 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 conditioned 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. To reveal potential mechanisms of this rich behavioral variability, we used a computational model to show that associative learning, in combination with a winner-takes-all decision process, can express a complex variability of behavior, similar to that observed in the data. The fact that the context can establish direct associations with specific responses is critical for our model's ability to account for the data.

Smooth and abrupt transitions upon the onset of extinction. Previous studies have reported that, after the onset of the extinction phase, the previously reinforced response can persist for several trials before showing signs of decay (39–41). This is indeed what we observed in 62% of the individually analyzed sessions (Figs. 2B1 and B3). However, in the remainder of the sessions, behavior changed abruptly at the onset of the extinction phase (Figs. 2B2 and B4). Remarkably, the type of transition was not evenly distributed across sessions: Smooth transitions dominated the first session, whereas abrupt transitions emerged exclusively after the first session (Fig. 3B). In previous studies, abrupt transitions might have been masked either because subjects were exposed to only one session (49, 50), or because results were pooled across several sessions/subjects (36), or they were not present because reinforcers were gradually reduced over time (51). In our model, the prevalence of smooth transitions during the first session, and the fact that abrupt transitions occurred almost exclusively in later sessions, is a consequence of the history of associations between the context and specific responses (17, 18, 21, 52, 53). 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 4C1), but once established, they can exert an effect on the behavior in later sessions (e.g. Fig. 4C2-4). 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.

What underlies the preference for the alternative choice during extinction? It has been shown that extinction is not limited to a decrease in a previously reinforced behavior, but it can also drive the emergence of new, previously non-reinforced behavior (23–27). In agreement with these studies, our results showed that during the extinction phase, the reinforced choice was not simply replaced by choice omissions. Instead, animals expressed a significant preference for the alternative choice in nearly one third of the sessions (Fig. 2B3 and B4), although this particular response was never reinforced in the presence of the discriminant stimulus in a given session. This counterintuitive behavior 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 (54, 55). 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.

Notwithstanding, 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 (56, 57), 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 dominated by omissions (30). 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. In our model, the emergence of persistent unrewarded behavior critically depends on the context's ability to establish excitatory and inhibitory associations with specific responses (17, 18, 21, 52, 53). This property enables the context to generate imbalances in the net inputs to motor units, allowing the emergence of persistent alternative choices during the extinction phase.

The renewal effect across sessions. To the best of our knowledge, the renewal effect has not been studied systematically across multiple sessions. We have shown here that the renewal effect was intermittent across sessions in single pigeons, but its overall prevalence across all pigeons decayed systematically as sessions progressed. In previous studies, this effect could have been masked by the pooling across sessions (36). In our model, the renewal effect occurs due to the release of the inhibition exerted by the context on a specific response (17–19, 21). After many sessions of training, these negative associations between context and responses reach their saturation value. As a consequence, the conditioned 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 (15, 47, 48, 58, 59).

The role of context in extinction and renewal. As discussed above, a key aspect of our model is that the context can directly establish excitatory and inhibitory associations with specific responses (17, 18, 21, 52, 53). However, it has been suggested that context can also establish associations with the representations of the outcome (60, 61), or modulate the association between response and outcome (occasion setting hypothesis) (62). 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. Nevertheless, our results do lend support to the idea that the context can establish direct excitatory and inhibitory associations with specific responses. If the effect of the context was limited to either a direct or modulatory influence over the representation of the outcome, 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.

Scope of the model. The purpose of our model was to test whether unexpected and highly variable individual behaviors can emerge from purely associative learning, i.e., without the influence of complex cognitive processes. Therefore, we have intentionally omitted components from previous associative models that could be attributed to higher-order cognitive functions. 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 (15, 47, 48, 58, 59). In some of these models, this term is updated from trial to trial, thereby modulating the learning rates dynamically within single sessions (48, 58, 59). 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.

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, particularly in species that easily learn abstract rules like corvids and primates (63, 64), in contrast to pigeons (42–44). It could be argued, for example, that the decay of the renewal effect with session number (Fig. 3D) reflects learning about the structure of the task: 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. S3, 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. 3B) 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. Based on our data, we cannot rule out that these hypothetical behavioral strategies drive the observed complex behaviors during extinction and renewal-test. However, based on the results of our simple associative learning model, we can conclude that higher-order cognition is not necessary to account for the aforementioned features in the cumulative learning curves.

In conclusion, we have uncovered a rich variability of behavior in operant 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.

Materials and Methods

Subjects. Twelve pigeons (*Columba livia*) 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 (36) situated in sound-attenuating cubicles (80cm x 80cm x 80cm)(65). Each Skinner box featured three rectangular pecking areas 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 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 (66).

Procedure. We employed a modified version of a consecutive extinction learning paradigm in which animals undergo an acquisition, extinction and a renewal-test phase within one session (67). The animals had to associate stimuli with corresponding choices. In the experiment, one single stimulus was presented per trial and signalled the animal to make either a left or a right choice at the end of the trial depending on the stimulus identity. 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. 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, a 2s long reward period commenced during which the food hopper was illuminated and food was available. In the case of an incorrect choice, the lights in the operant-chamber were turned off for 2s as a mild punishment. Consecutive trials were separated by an inter-trial-interval (ITI) of 4s duration. The structure of the trials for the different experimental phases is shown in Fig. 1A. 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. In a trial, only one of the four stimuli was presented on the center key. Animals were pre-trained on two of the stimuli prior to the experimental sessions studied here. 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 associations had to be learned in the acquisition phase through trial-and-error.

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. The subsequent extinction phase was marked by two key differences as compared to the acquisition phase. (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. 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 conditioned 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. Finally, the renewal-test phase was used to study the return of the conditioned 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.

Behavioral analysis. 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. For each trial, the choice behavior of the animal, namely, omissions, alternative choice, and conditioned choice, were encoded as 0, -1, and 1, respectively. To quantify the preference for the alternative and conditioned choices during the extinction and renewal-test phases, respectively, we counted the number (k) of responses and assessed its significance by calculating the probability $p(k)$ of obtaining at least k responses in N random trials under the null hypothesis. N is the total number of trials, in which a particular stimulus was presented in a particular phase. 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 can occur with probability ⅓. If the probability of observing at least k responses 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. 2A and S1), 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 method to test for the renewal effect, which was regarded as present when the animals significantly expressed the conditioned choice during the renewal-test phase. Due to the relatively small number of animals (7) that underwent more than 10 sessions, to correlate the session number with the prevalence of the renewal effect across animals, we grouped the data from sessions 11 to 22 in blocks of two sessions (10 to 13 data points per block), and the data from sessions 23 to 26 in one single block (13 data points). 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 conditioned 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. 3A). 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 Δw_{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 (15, 47, 48, 58, 59).

Acknowledgments

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

References

1. B. F. Skinner, *The behavior of organisms: an experimental analysis* (Appleton-Century, 1938).
2. M. E. Bouton, Learning and the persistence of appetite: Extinction and the motivation to eat and overeat. *Physiol. Behav.* **103**, 51–58 (2011).
3. B. J. Everitt, T. W. Robbins, Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nat. Neurosci.* **8**, 1481–1489 (2005).
4. S. E. Hyman, Addiction: A Disease of Learning and Memory. *Am. J. Psychiatry* **162**, 1414–1422 (2005).
5. A. E. Kelley, Memory and Addiction: Shared Neural Circuitry and Molecular Mechanisms. *Neuron* **44**, 161–179 (2004).
6. A. D. Redish, Addiction as a Computational Process Gone Awry. *Science* **306**, 1944–1947 (2004).
7. M. Barad, P.-W. Gean, B. Lutz, The Role of the Amygdala in the Extinction of Conditioned Fear. *Biol. Psychiatry* **60**, 322–328 (2006).
8. I. R. Galatzer-Levy, G. A. Bonanno, D. E. Bush, J. LeDoux, Heterogeneity in threat extinction learning: substantive and methodological considerations for identifying individual difference in response to stress. *Front. Behav. Neurosci.* **7** (2013).
9. J. A. Gottfried, R. J. Dolan, Human orbitofrontal cortex mediates extinction learning while accessing conditioned representations of value. *Nat. Neurosci.* **7**, 1144–1152 (2004).
10. M. R. Milad, *et al.*, Recall of Fear Extinction in Humans Activates the Ventromedial Prefrontal Cortex and Hippocampus in Concert. *Biol. Psychiatry* **62**, 446–454 (2007).
11. D. Lengersdorf, M. C. Stüttgen, M. Uengoer, O. Güntürkün, 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 (2014).
12. M. Gao, D. Lengersdorf, M. C. Stüttgen, O. Güntürkün, NMDA receptors in the avian amygdala and the premotor arcopallium mediate distinct aspects of appetitive extinction learning. *Behav. Brain Res.* **343**, 71–82 (2018).
13. M. Eisenberg, Y. Dudai, Reconsolidation of fresh, remote, and extinguished fear memory in medaka: old fears don't die. *Eur. J. Neurosci.* **20**, 3397–3403 (2004).
14. N. Stollhoff, R. Menzel, D. Eisenhardt, 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 (2005).
15. R. A. Rescorla, A. R. Wagner, "A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement." in *Classical Conditioning II: Current Research and Theory*, A. H. Black, W. F. Prokasy, Eds. (Appleton-Century-Crofts, 1972), pp. 64–99.
16. J. L. McClelland, D. E. Rumelhart, Distributed memory and the representation of general and specific information. *J. Exp. Psychol. Gen.* **114**, 159–197 (1985).

17. M. E. Bouton, Extinction of instrumental (operant) learning: interference, varieties of context, and mechanisms of contextual control. *Psychopharmacology (Berl.)* **236**, 7–19 (2019).
18. T. P. Todd, Mechanisms of renewal after the extinction of instrumental behavior. *J. Exp. Psychol. Anim. Behav. Process.* **39**, 193–207 (2013).
19. S. Trask, E. A. Thrailkill, M. E. Bouton, Occasion setting, inhibition, and the contextual control of extinction in Pavlovian and instrumental (operant) learning. *Behav. Processes* **137**, 64–72 (2017).
20. E. Tulving, D. M. Thomson, Encoding specificity and retrieval processes in episodic memory. *Psychol. Rev.* **80**, 352–373 (1973).
21. T. P. Todd, D. Vurbic, M. E. Bouton, Mechanisms of renewal after the extinction of discriminated operant behavior. *J. Exp. Psychol. Anim. Learn. Cogn.* **40**, 355–368 (2014).
22. C. A. Conklin, S. T. Tiffany, Applying extinction research and theory to cue-exposure addiction treatments. *Addiction* **97**, 155–167 (2002).
23. P. R. Fuller, Operant Conditioning of a Vegetative Human Organism. *Am. J. Psychol.* **62**, 587–590 (1949).
24. J. J. Antonitis, Response variability in the white rat during conditioning, extinction, and reconditioning. *J. Exp. Psychol.* **42**, 273–281 (1951).
25. M. R. Tinsley, W. Timberlake, M. Sitomer, D. R. Widman, 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 (2002).
26. M. K. Lattal, K. A. Lattal, Facets of Pavlovian and operant extinction. *Behav. Process.* **90**, 1–8 (2012).
27. L. L. Grow, M. E. Kelley, H. S. Roane, M. A. Shillingsburg, Utility of Extinction-Induced Response Variability for the Selection of Mands. *J. Appl. Behav. Anal.* **41**, 15–24 (2008).
28. M. A. E. André, O. Güntürkün, D. Manahan-Vaughan, The metabotropic glutamate receptor, mGlu5, is required for extinction learning that occurs in the absence of a context change. *Hippocampus* **25**, 149–158 (2015).
29. M. Méndez-Couz, J. M. Becker, D. Manahan-Vaughan, Functional Compartmentalization of the Contribution of Hippocampal Subfields to Context-Dependent Extinction Learning. *Front. Behav. Neurosci.* **13** (2019).
30. A. D. Redish, S. Jensen, A. Johnson, Z. Kurth-Nelson, Reconciling reinforcement learning models with behavioral extinction and renewal: Implications for addiction, relapse, and problem gambling. *Psychol. Rev.* **114**, 784–805 (2007).
31. S. Lissek, B. Glaubitz, O. Güntürkün, M. Tegenthoff, Noradrenergic stimulation modulates activation of extinction-related brain regions and enhances contextual extinction learning without affecting renewal. *Front. Behav. Neurosci.* **9**, 34 (2015).
32. J. A. Larrauri, N. A. Schmajuk, Attentional, associative, and configural mechanisms in extinction. *Psychol. Rev.* **115**, 640–676 (2008).
33. S. J. Gershman, D. M. Blei, Y. Niv, Context, learning, and extinction. *Psychol. Rev.* **117**, 197–209 (2010).
34. S. Wirth, *et al.*, Single neurons in the monkey hippocampus and learning of new associations. *Science* **300**, 1578–81 (2003).
35. S. Wirth, *et al.*, Trial Outcome and Associative Learning Signals in the Monkey Hippocampus. *Neuron* **61**, 930–940 (2009).
36. J. Packheiser, O. Güntürkün, R. Pusch, 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 (2019).
37. C. R. Gallistel, S. Fairhurst, P. Balsam, The learning curve: Implications of a quantitative analysis. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 13124–13131 (2004).
38. J. C. Leslie, *et al.*, Effects of Reinforcement Schedule on Facilitation of Operant Extinction by Chlordiazepoxide. *J. Exp. Anal. Behav.* **84**, 327–338 (2005).
39. C. A. Podlesnik, T. A. Shahan, Behavioral momentum and relapse of extinguished operant responding. *Learn. Behav.* **37**, 357–364 (2009).
40. C. A. Podlesnik, T. A. Shahan, Extinction, relapse, and behavioral momentum. *Behav. Processes* **84**, 400–411 (2010).
41. J. A. Nevin, Resistance to extinction and behavioral momentum. *Behav. Processes* **90**, 89–97 (2012).
42. E. Maes, *et al.*, Feature- versus rule-based generalization in rats, pigeons and humans. *Anim. Cogn.* **18**, 1267–1284 (2015).
43. J. D. Smith, *et al.*, Pigeons' categorization may be exclusively nonanalytic. *Psychon. Bull. Rev.* **18**, 414–421 (2011).
44. B. Wilson, N. J. Mackintosh, R. A. Boakes, Transfer of relational rules in matching and oddity learning by pigeons and corvids. *Q. J. Exp. Psychol. Sect. B* **37**, 313–332 (1985).
45. A. Ghazizadeh, F. Ambroggi, N. Odean, H. L. Fields, Prefrontal Cortex Mediates Extinction of Responding by Two Distinct Neural Mechanisms in Accumbens Shell. *J. Neurosci.* **32**, 726–737 (2012).
46. J. Felsenberg, *et al.*, Integration of Parallel Opposing Memories Underlies Memory Extinction. *Cell* **175**, 709-722.e15 (2018).
47. R. R. Bush, F. Mosteller, A mathematical model for simple learning. *Psychol. Rev.* **58**, 313–323 (1951).
48. N. J. Mackintosh, A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychol. Rev.* **82**, 276–298 (1975).
49. V. L. Kinner, C. J. Merz, S. Lissek, O. T. Wolf, Cortisol disrupts the neural correlates of extinction recall. *NeuroImage* **133**, 233–243 (2016).
50. S. Lissek, B. Glaubitz, M. Uengoer, M. Tegenthoff, Hippocampal activation during extinction learning predicts occurrence of the renewal effect in extinction recall. *NeuroImage* **81**, 131–43 (2013).
51. J. A. Nevin, R. C. Grace, Behavioral momentum and the Law of Effect. *Behav. Brain Sci.* **23**, 73–90 (2000).
52. R. A. Rescorla, Inhibitory associations between S and R in extinction. *Anim. Learn. Behav.* **21**, 327–336 (1993).
53. R. A. Rescorla, Response Inhibition in Extinction. *Q. J. Exp. Psychol. Sect. B* **50**, 238–252 (1997).
54. H. Leitenberg, R. A. Rawson, K. Bath, Reinforcement of Competing Behavior during Extinction. *Science* **169**, 301–303 (1970).
55. N. E. Winterbauer, M. E. Bouton, Mechanisms of resurgence of an extinguished instrumental behavior. *J. Exp. Psychol. Anim. Behav. Process.* **36**, 343–353 (2010).
56. A. Gharib, C. Gade, S. Roberts, Control of Variation by Reward Probability. *J. Exp. Psychol. Anim. Behav. Process.* **30**, 271–282 (2004).
57. P. Anselme, O. Güntürkün, How foraging works: Uncertainty magnifies food-seeking motivation. *Behav. Brain Sci.* **42** (2019).

58. J. M. Pearce, G. Hall, A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol. Rev.* **87**, 532–552 (1980).
59. N. A. Schmajuk, J. W. Moore, Real-time attentional models for classical conditioning and the hippocampus. *Physiol. Psychol.* **13**, 278–290 (1985).
60. A. G. Baker, H. Steinwald, M. E. Bouton, Contextual conditioning and reinstatement of extinguished instrumental responding. *Q. J. Exp. Psychol. Sect. B* **43**, 199–218 (1991).
61. J. M. Pearce, G. Hall, The influence of context-reinforcer associations on instrumental performance. *Anim. Learn. Behav.* **7**, 504–508 (1979).
62. S. Trask, M. E. Bouton, Contextual control of operant behavior: evidence for hierarchical associations in instrumental learning. *Learn. Behav.* **42**, 281–288 (2014).
63. L. Veit, A. Nieder, Abstract rule neurons in the endbrain support intelligent behaviour in corvid songbirds. *Nat. Commun.* **4**, 1–11 (2013).
64. I. M. White, S. P. Wise, Rule-dependent neuronal activity in the prefrontal cortex. *Exp. Brain Res.* **126**, 315–335 (1999).
65. J. Packheiser, *et al.*, How competitive is cue competition? *Q. J. Exp. Psychol.* **73**, 104–114 (2020).
66. J. Rose, T. Otto, L. Dittrich, The Biopsychology-Toolbox: A free, open-source Matlab-toolbox for the control of behavioral experiments. *J. Neurosci. Methods* **175**, 104–107 (2008).
67. S. Starosta, M. C. Stüttgen, O. Güntürkün, Recording single neurons' action potentials from freely moving pigeons across three stages of learning. *J Vis Exp* **88**, 51283 (2014).

Figures and Tables

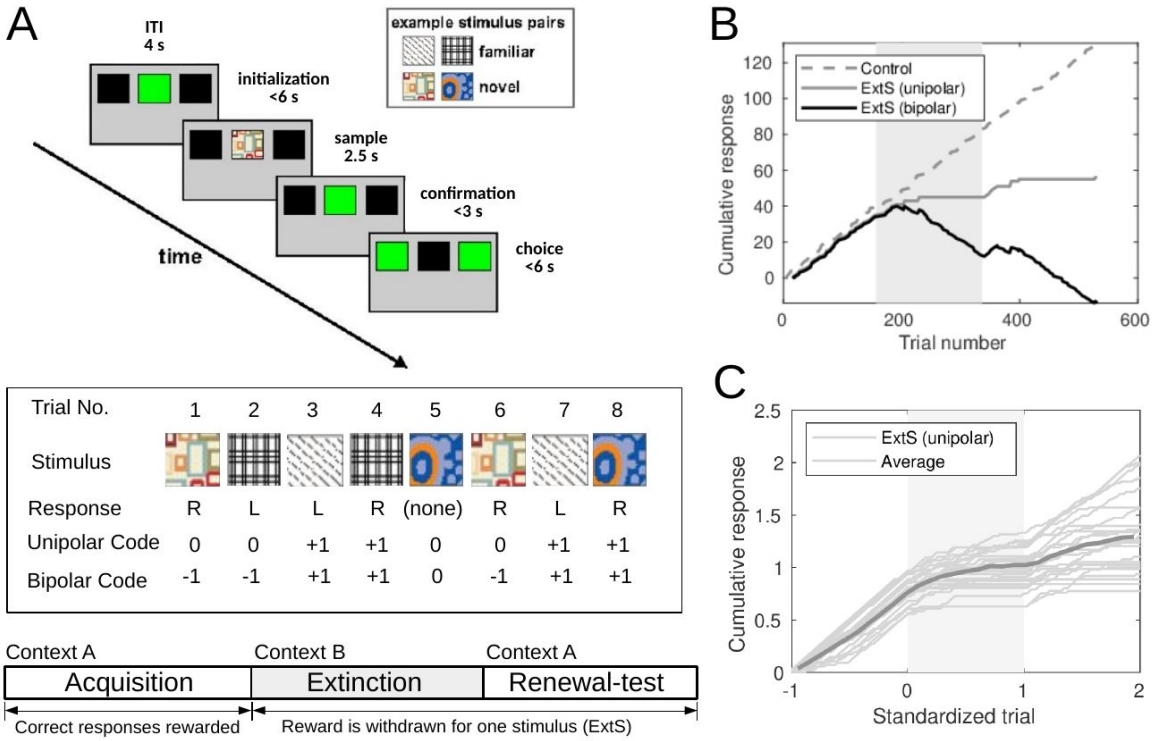

Figure 1. *An operant extinction task exhibits complex choice behavior across sessions.* (A) Illustration of the behavioral paradigm. Top: Each trial is initialized by a peck on the central key, which triggers the presentation of one of four stimuli. After stimulus presentation, animals confirm with a peck on the central key, and choose between either pecking the left or right key, or omitting the response altogether. Middle: The trial outcomes, i.e. responses ("R" stands for right peck, "L" for left peck and "(none)" for omission), can be encoded in a unipolar (0,1) or bipolar fashion (-1,0,+1). Bottom: Each session consisted of three phases with different reward contingencies for the extinction stimulus (ExtS). (B) Learning curves obtained from one session. Unipolar coding (solid gray trace) shows the decay of the conditioned response in the extinction phase (gray area) and the renewal of the conditioned 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. (C) 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)).

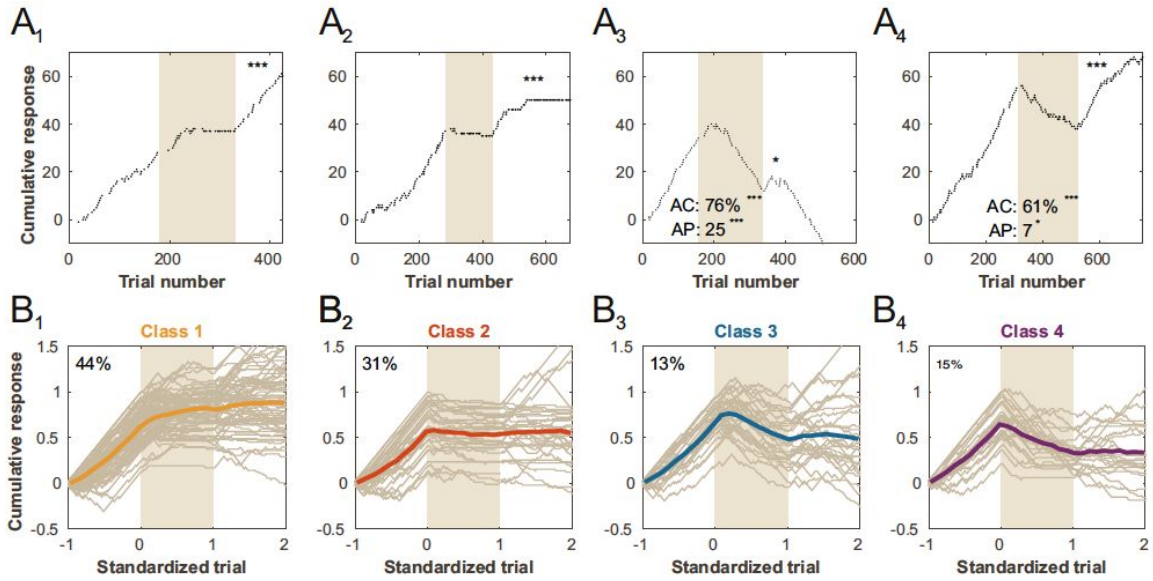

Figure 2. *Variability of behavior during extinction.* (A) Bipolar-encoded cumulative learning curves for a sample of four sessions. Behavior during the extinction phase is not limited to the canonical extinction curve (A1), which is characterized by smooth transitions upon the onset of context B, and extinction dominated by omissions. Learning curves also express abrupt transitions (A2 and 4) and preference for the alternative choice (A3 and 4). Proportion of alternative-choice responses (AC) and persistence on alternative (AP; number of consecutive trials during the respective phase) are shown when significant ($p < 0.05$). Stars on top of the learning curves mark the significance of the renewal effect. (B) Standardized learning curves (gray traces) corresponding to 156 behavioral sessions obtained from 12 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. B1: smooth transition and no alternative choice; B2: abrupt transition and no alternative choice. B3: smooth transition and alternative choice. B4: 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.

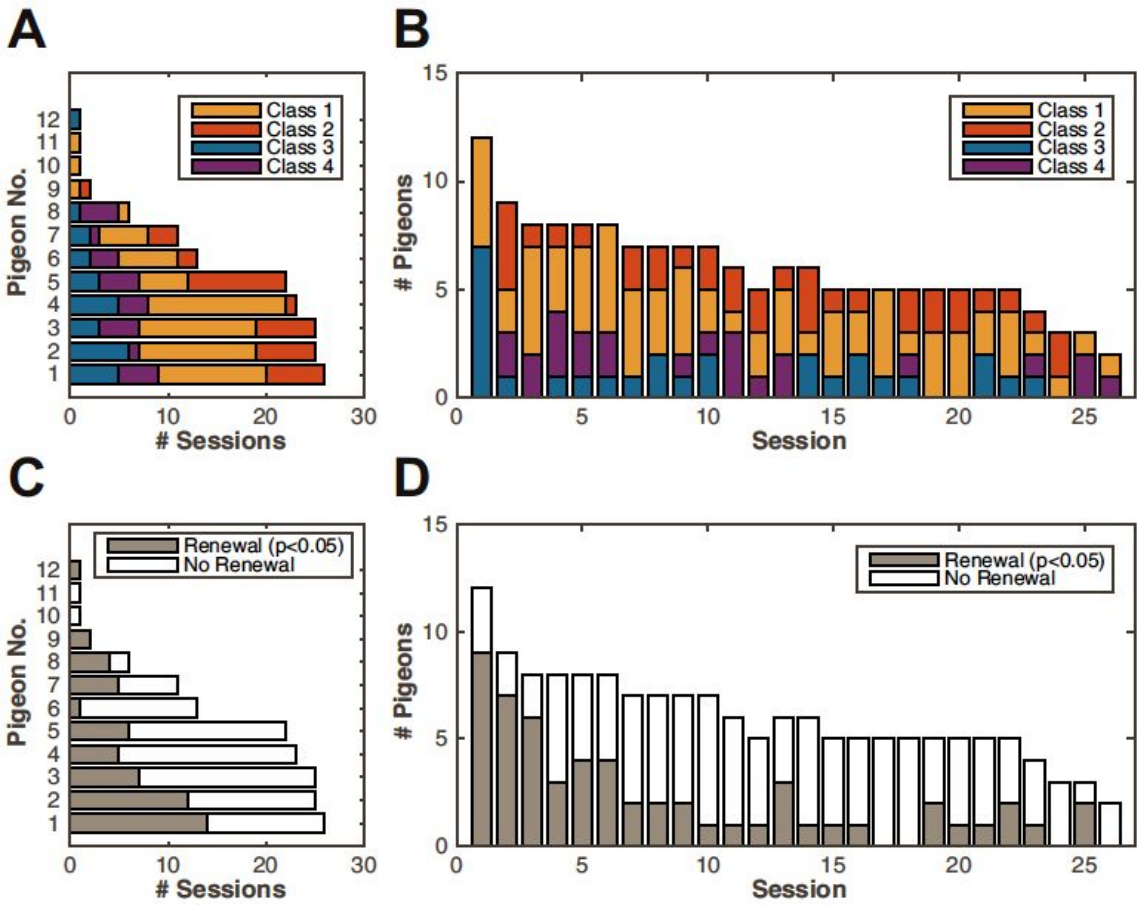

Figure 3. *Variability of behavior across pigeons and sessions.* (A) Number of learning curves (# sessions) that fall into each of the four classes for each animal. Individual pigeons do not exhibit a clear bias for a particular type of learning curve. (B) Number of animals expressing each class of learning 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. (D) Expression of the renewal effect declines as a function of session.

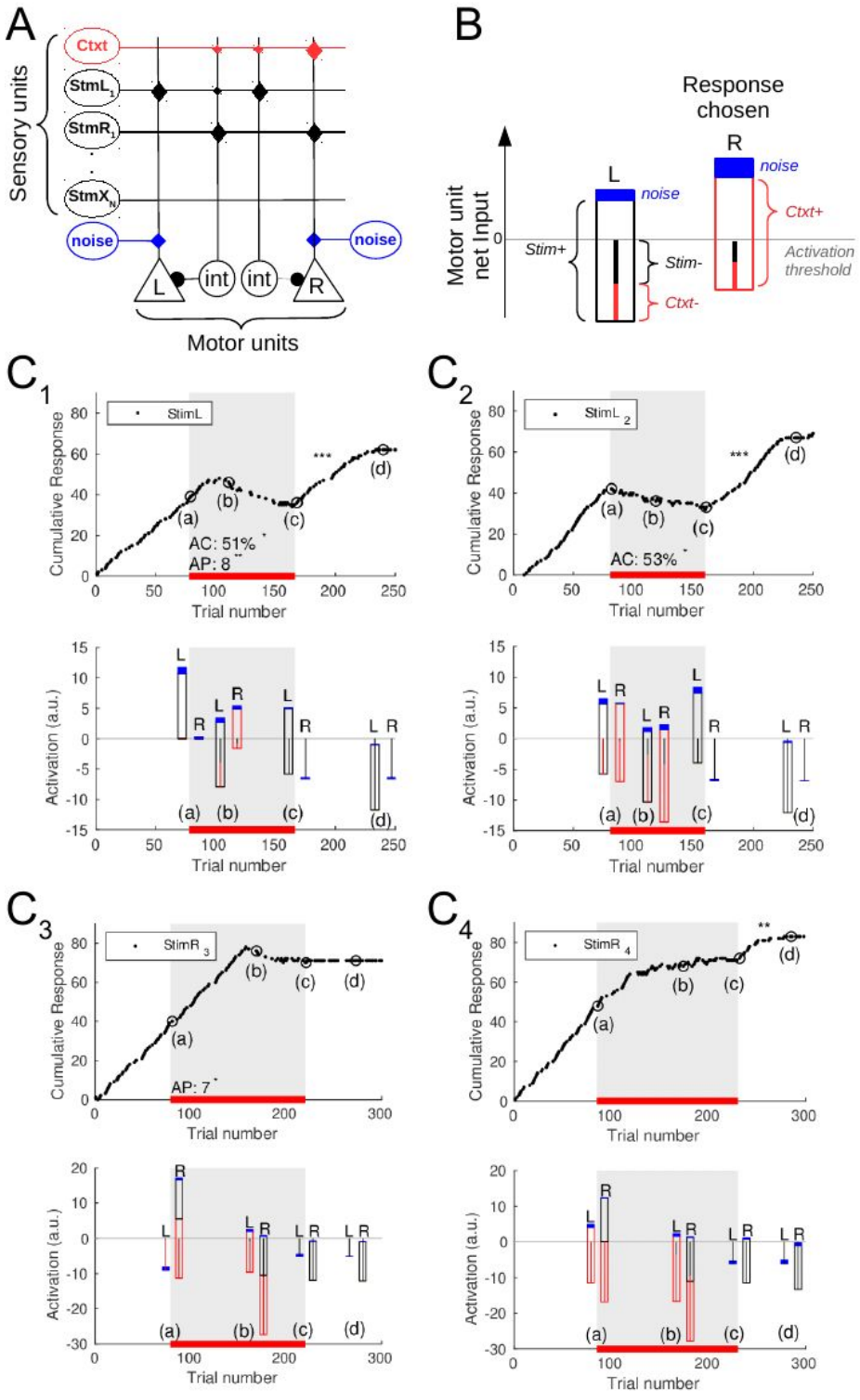

Figure 4. *Associative learning model predicts extinction dominated by alternative choice.* (A) Associative network. Sensory units (ovals) can establish excitatory associations directly with motor units (triangles) mediating the left and right responses, or inhibitory associations via interneurons (circles). 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).

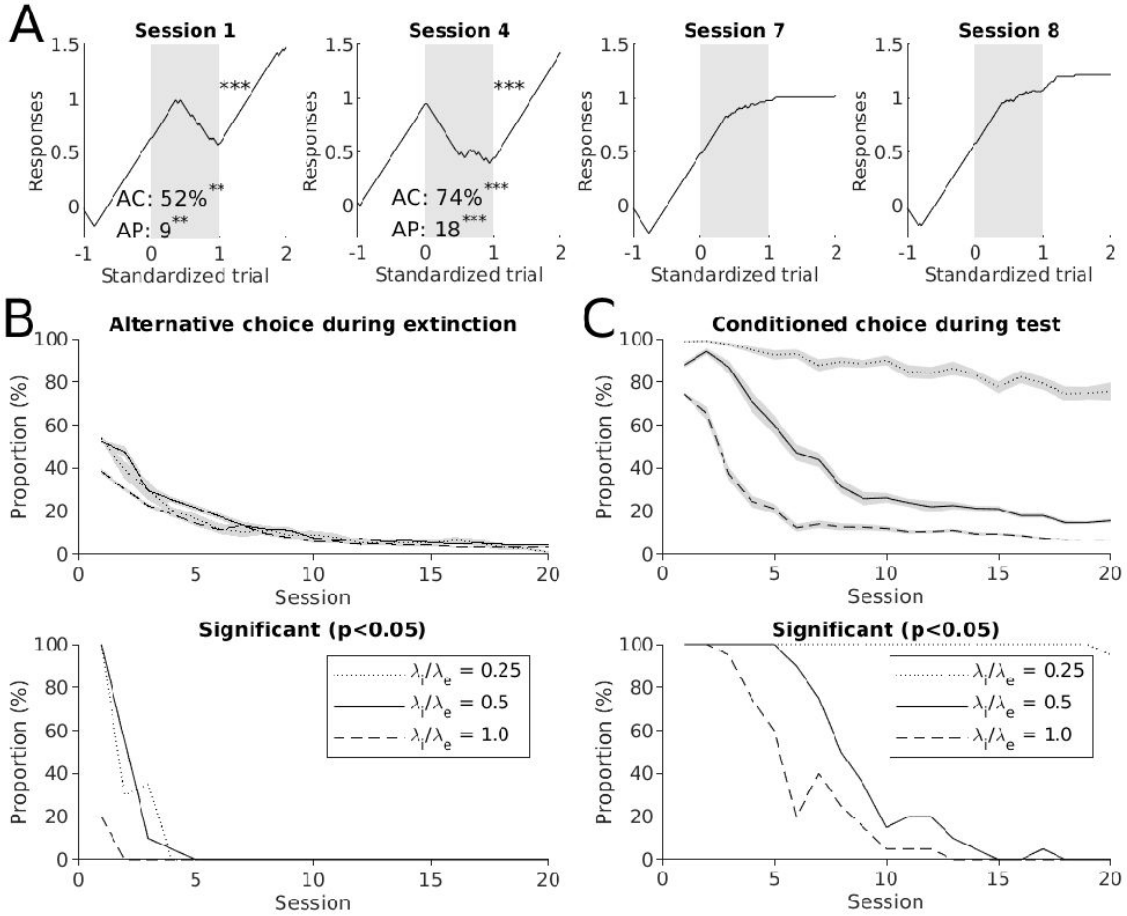

Figure 5. *Associative learning accounts for the general trend observed in the behavior of pigeons in extinction learning and renewal.* (A) Sample sessions obtained from one pigeon-model, showing strong preference for the alternative choice (sessions 1 and 4), an abrupt transition at the onset of the 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 conditioned choices (top) and proportion of pigeon-models emitting a significant number of conditioned 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

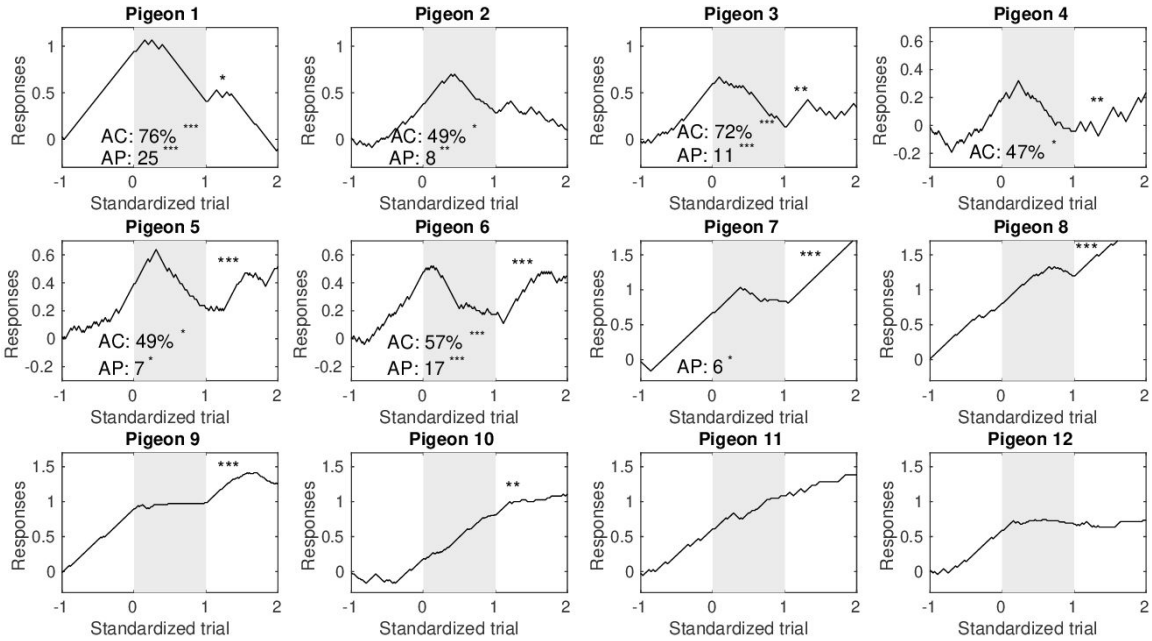

Figure S1. *Standardized learning curves from the first exposure to the extinction task.* Gray area demarcates the extinction phase. Proportion of alternative choices (AC; percent during the respective phase) and alternative persistence (AP; number of consecutive trials during the respective phase) are shown when significant ($p < 0.05$). Stars on top right of panels mark the significance of the renewal effect.

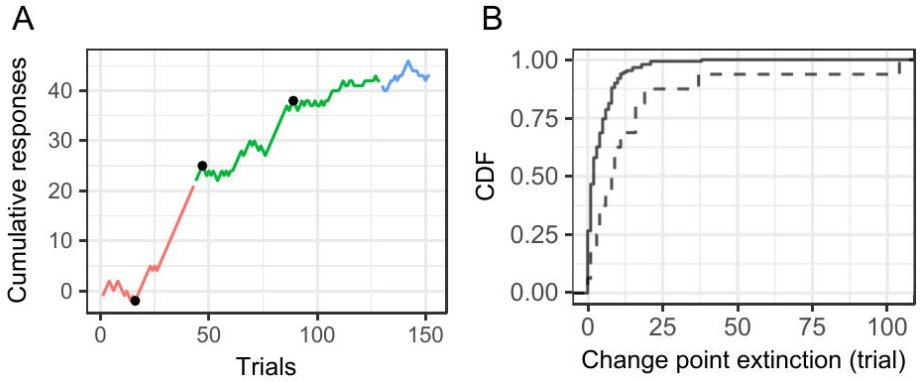

Figure S2. *Change point analysis reveals changes of behavior across sessions*. (A) Cumulative responses to the extinction stimulus in a single session. Phase is color-coded (red: acquisition; green: extinction; blue: renewal-test). Black dots show the change-points as identified by the change point 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.

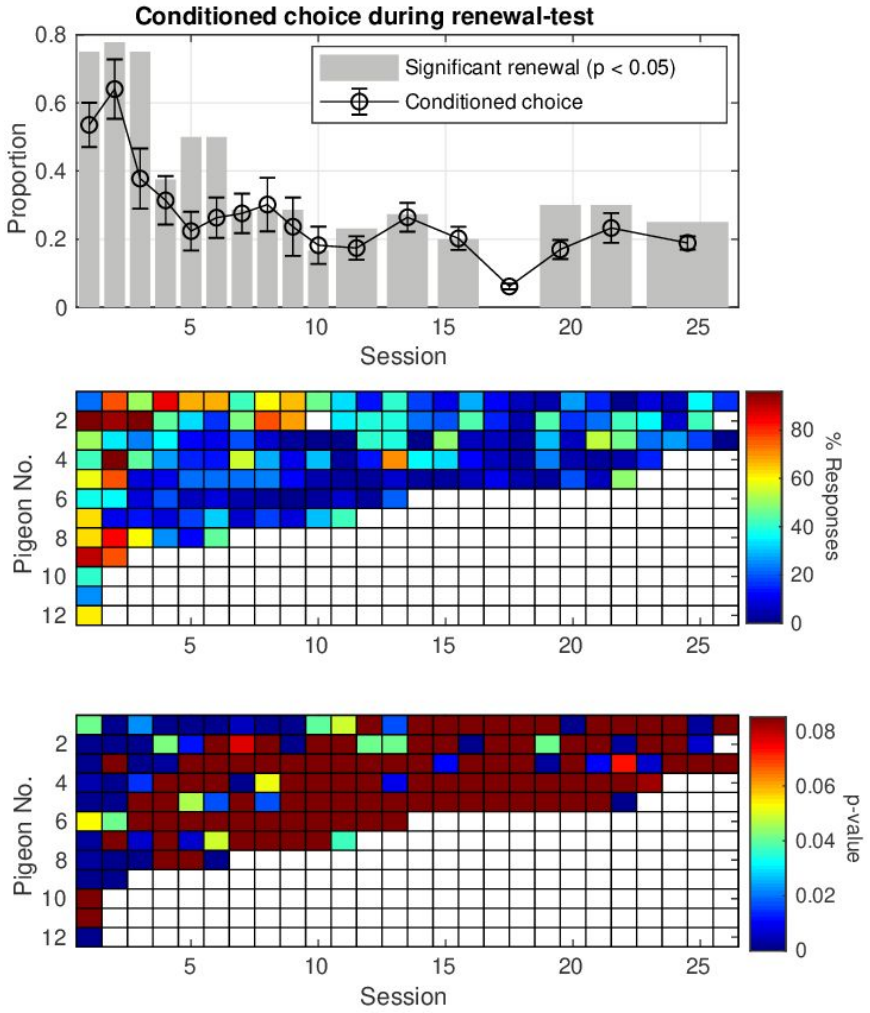

Figure S3. *Distribution of the renewal effect across pigeons and sessions.* Top: Proportion of pigeons expressing a significant number of conditioned choices (gray bars) and average fraction of conditioned choices (black line) during renewal-test as a function of session block. Due to the relatively small number of animals (7) that underwent more than 10 sessions, data from sessions 11 to 22 were grouped in blocks of two sessions (10 to 13 data points per block), and data from sessions 23 to 26 were grouped in one single block (13 data points) Middle: Distribution of occurence of conditioned choices for all pigeons and sessions. Bottom: p-values of the conditioned choices for all pigeons and sessions.