Review article

Onur Güntürkün*, Maik C. Stüttgen, Sarah Starosta, Roland Pusch, Meng Gao, Michael Nitsche, Thomas M. Ernst, Mark E. Ladd, Harald H. Quick and Dagmar Timmann

Beyond the classic extinction network: a wider, comparative view

https://doi.org/10.1515/nf-2020-0015

Abstract: Extinction learning modifies the dynamics of brain circuits such that a previously learned conditioned response is no longer generated. The majority of extinction studies use fear conditioning in rodents and identified the prefrontal cortex, the hippocampus, and the amygdala as core regions of the extinction circuit. We sought to find answers to two questions: First, do we find a similar functional brain circuit in birds, which underwent a 300-million-year separate evolution from mammals? Second, do we have to incorporate the cerebellum as a key component of the central extinction circuit? We indeed show that the avian extinction pathways are not identical but highly similar to those of mammals. In addition, we reveal that the human cerebellum processes prediction errors, a key element driving extinction of learned fear responses, and contributes to context-related effects of extinction.

Keywords: cerebellum; context learning; eyeblink conditioning; pigeons; prediction error.


Schlüsselwörter: Kleinhirn; Kontextlernen; Lidenschlagkonditionierung; Tauben; Vorhersagefehler.

Introduction

Animals rapidly learn to predict which stimuli are followed by reward or punishment, or, in more general terms, by an expected unconditioned stimulus (US). Conversely, this learned association can change when the US is omitted after stimulus presentation. This latter process is known as extinction learning and constitutes one of the most fundamental learning mechanisms (Rescorla and Wagner, 1972). Decades of research show that extinction of a conditioned response due to withholding of the US does...
not merely involve forgetting the original association but entails new learning. The principles of extinction learning were demonstrated to be largely similar in animals that reach from humans (Icenhour et al., 2015) to insects (Felsenberg et al., 2018). If a learning “law” occurs in so diverse species with similar or even identical mechanisms, we should expect overlapping neural processes of extinction learning across the animal kingdom. But is this indeed the case? This question is at the core of the first part of this article in which we study extinction circuits in pigeons. Because birds have a more-than-300-million-year-old separate evolutionary history from mammals, we can test, if, at least among amniotes, the neural fundaments of extinction learning are invariant.

We then move on to the extinction-relevant pathways in mammals. Using fear extinction paradigms in rodents, three key neural structures were identified to be at the core of this system: the amygdala, prefrontal cortex, and hippocampus (Orsini and Maren, 2012). In the second part of our article, we aim to add the cerebellum as an overlooked, but, in our opinion, important structure to the established extinction circuitry. As we will show, the cerebellum plays an important role in the processing of prediction errors in sensory and reward-related domains and thus controls core elements of associative learning.

The avian neural circuit for extinction learning

To identify the neural circuit for extinction learning in an appetitive paradigm, we trained pigeons in a within-subject renewal design to peck on two conditioned stimuli (CSs) in two different contexts (sign tracking). Immediately before an extinction session, animals received intracranial injections of saline or drug. In different studies, we used either AP5 to inhibit local N-methyl-D-aspartate (NMDA) receptors or tetrodotoxin (TTX) to block Na+ channels. This intervention was followed by extinction training in the opposite context. Subsequently, pigeons were tested for retrieval of extinction memory in both contexts.

As depicted in Figure 1, visual information about conditional cues ascends via visual pathways to the visual associative nidopallium frontolaterale (NFL). TTX injections into NFL slow down extinction learning and reduce retrieval of context-specific extinction information. Notably, this effect is not due to perceptual impairment during learning (Gao et al., 2019a). Multiple projections fan out of NFL, and one of them leads to the hippocampus. Here, TTX injections caused no deficits in extinction learning but affected the consolidation of extinction memory. Importantly, we obtained no strong evidence for a hippocampus-mediated context dependency of extinction memory (Lengersdorf et al., 2014), in contrast to findings in rodents (Maren and Hobin, 2007). NFL also projects to the nidopallium caudolaterale (NCL) – the avian functional equivalent to the prefrontal cortex. Transiently inactivating the NCL with TTX did not affect extinction learning but impaired consolidation of extinction memory (Lengersdorf et al., 2014). In addition, multiple studies indicate a role of the avian NCL in the integration of context information into extinction memory (Lissek and Güntürkün, 2005; Starosta et al., 2017). NFL also projects to the medial striatum, the NCL, the avian amygdala, and the arcopallium. This last structure is the avian analog to the pre/motor cortex. Inhibiting NMDA receptors in the medial striatum or the amygdala impairs extinction learning, while the same procedure impairs consolidation of extinction memory in the arcopallium (Gao et al., 2018, 2019b; Lengersdorf et al., 2015).

Taken together, the visual-associative NFL, prefrontal-like NCL, amygdala, and medial striatum (StM) are involved in extinction learning. Our pharmacological treatment in these areas possibly impaired the updating of reward prediction errors and thus caused the deficit in extinction learning. A further cluster of structures (hippocampus, NCL, pre/motor arcopallium) is required for...
consolidation of extinction memory. Finally, NFL and NCL play important roles in the context modulation of extinction learning.

By and large, this pattern strongly resembles the systems architecture of extinction learning in mammals (Milad and Quirk, 2012). This would speak in favor of an ancient functional forebrain architecture that goes back more than 300 million years. There is, however, one glaring difference: While a key function of the mammalian hippocampus is the processing of context-dependent extinction information, we found not much evidence for this in pigeons. It is conceivable that the lack of direct connectivity between the avian “prefrontal” NCL and hippocampus is the key difference that drives this functional dissimilarity. This functional characterization of our current understanding of the extinction circuit in pigeons will pave the way for deeper functional analyses of these areas during extinction learning. This is shown for the NCL in the next part.

**Single neurons in the avian forebrain dynamically encode acquired and extinguished associations**

To elucidate the neuronal underpinnings of extinction learning, we recorded from single neurons in the prefrontal-like NCL during learning. An ideal paradigm to investigate extinction learning allows the observation of single-neuron activity during not only the extinction of conditioned responding but also the preceding acquisition and the subsequent reappearance of responding (spontaneous recovery or reacquisition). To this end, we designed a task that encompasses these three stages of learning in a single behavioral session (Starosta et al., 2014). Single-neuron activity was recorded while animals acquired an instrumental response to a novel visual stimulus for reward, which was subsequently extinguished (reward omission: extinction) and then reestablished (reward for responding was reintroduced: reacquisition). In our task, pigeons were confronted with one of several visual stimuli on one response key and had to learn which of the two adjacent choice keys was associated with that stimulus (Figure 2A). Two stimuli were familiar to the animals from earlier sessions, while two others were new such that the correct response had to be learned. After reaching learning criterion, the response to one of the new stimuli was no longer reinforced (extinction). Once the performance for this stimulus dropped below 65%, the response was reinforced again (reacquisition). Figure 2B shows behavioral results from an example session. Figure 2C–F summarizes the behavioral results of five animals performing this task repeatedly. It illustrates that acquisition, extinction, and reacquisition phases are associated with an increase, decrease, and second increase in performance, respectively. As expected, extinction leads to a decrease of pecks onto the visual stimulus (Figure 2D) and an increase in reaction times which was reversed during reacquisition (Figure 2E). Finally, the number of trials to criterion performance is higher in acquisition than in reacquisition, in line with the hypothesis that extinction is a new learning process and not mere forgetting (Figure 2F).

On a neuronal level, we reasoned that learning affects neuronal responses such that activity should change across the three learning stages (Veit et al., 2015). This was indeed the case: Figure 2G shows neurometric curves from an NCL neuron whose activity profile changed in the course of learning. Specifically, the neuron discriminated the two familiar stimuli (blue curves) almost perfectly across the entire session (area under the receiver operating characteristic curve, AUROC: 1, perfect discriminability; 0.5, no discriminability). In contrast, discriminability for the two novel stimuli changed dramatically over the course of the session: during acquisition, discriminability was relatively constant but moderate; in extinction, discriminability decreased from nearly optimal to chance levels; during reacquisition, neural discriminability again increased.

This pattern was also seen in the population of NCL neurons. Figure 2H depicts in black neural discriminability for the two novel stimuli across the three stages of learning, separately averaged across selective NCL neurons ($n = 32$ [acquisition phase], 29 [extinction phase], 14 [reacquisition phase]) or not selective ($n = 187, 166, 127$). Familiar stimuli are shown in gray. Notably, only those neurons that were selective for the familiar stimuli (black curves) showed learning-related modulation. Thus, many NCL neurons reflected the strength of conditioned responding across learning stages. Taken together, our novel paradigm highlights diverse reorganization patterns of neuronal activity in single NCL units during learning. While a subpopulation of neurons faithfully tracks the “ups and downs” of associative strength, others seem to code further aspects of the task that could explain the saving of associative memory across extinction learning. These aspects will be uncovered by a deeper analysis of the activity patterns of these neurons. But only integrating these insights into the framework of the overall functional extinction circuits as outlined in the beginning will allow a deeper understanding of the neuronal dynamics during extinction.
extinction. This is exemplified in the next part of our article that demonstrates that the cerebellum, a hitherto neglected part of the extinction network, is in fact an important component of this system.

The cerebellum as a frequently ignored component of the extinction network

Comparatively little is known about the contribution of the cerebellum to extinction of learned fear responses (Apps and Strata, 2015). Cerebellar contribution to extinction has been studied in most detail in eyeblink conditioning (Hu et al., 2015, for reviews). As yet, most studies focused on the intrinsic cerebellar mechanisms involved in extinction but neglected additional cerebello-cerebral interactions. In the rodent literature, there is some evidence that learning-related changes of Purkinje cell activities in the cerebellar cortex are reversed during extinction. Recording studies show that Purkinje cells learn to reduce their activity (“pause”) in response to the CS during acquisition of conditioned eyeblink responses. This pause is reversed during extinction and returns during reacquisition (Jir恩 hed et al., 2007). The inhibitory feedback connection between the cerebellar nuclei and the inferior olive seems to play a critical role in extinction (Bengtsson et al., 2007; Medina et al., 2002). The results of our functional magnetic resonance imaging (fMRI) studies in humans agree with the hypothesis that at least parts of the initial learning memory in the cerebellar cortex are erased during extinction (Medina et al., 2002).

We established a setup that allows ultra-high-field 7T fMRI of the cerebellum during eyeblink conditioning in humans (Thürling et al., 2015). We found that activations in the cerebellar cortex related to the acquisition of conditioned eyeblink responses were reversed during extinction (Thürling et al., 2015). Findings were largely confirmed in a subsequent 7T fMRI study using the same setup in a different group of participants (Ernst et al., 2017). We were unable to show saving-related cerebellar activation (Ernst et al., 2017). These findings, however, do not exclude the possibility that parts of the initial memory trace remain in the cerebellum during extinction. The cerebellar nuclei, but also extracerebellar regions, may be potential
substrates of saving effects (Medina et al., 2001). But also extracerebellar regions may play a role (Kalmbach and Mauk, 2012). These regions may be under the inhibitory control of the known cerebral fear extinction network (Hu et al., 2015, for review), but this has been studied in detail neither in humans nor in animals.

Bidirectional learning within the cerebellar cortex implies that cortical areas involved in acquisition and extinction of learned associations at least partially overlap. Our findings in patients with cerebellar lesions agree with this assumption. We tested acquisition and extinction of conditioned eyeblink responses (Ernst et al., 2016) and acquisition and extinction of cognitive associations (Steiner et al., 2020) in patients with cerebellar disease. Patients who had preserved acquisition—a prerequisite to study extinction effects—showed extinction not different from controls.

Extinction, however, is known to be more context-dependent than acquisition and to involve a more extended cortical network, including the prefrontal cortex and the hippocampus (Milad and Quirk, 2012). Likewise, cerebellar areas involved in extinction may be more extended than cerebellar areas involved in acquisition. Initial findings in cerebellar patients support this assumption (Steiner et al., 2019). We studied patients with focal cerebellar disease and preserved acquisition of conditioned eyeblink responses. Extinction was not different from controls. Renewal effects, however, appeared to be impeded in patients with lesions of the more posterolateral cerebellar hemisphere which has connections with the prefrontal cortex and hippocampus (Bostan et al., 2013; Watson et al., 2019). Furthermore, we found activation of the posterolateral cerebellar hemisphere related to context change during extinction learning of cognitive associations in healthy participants in a 3T fMRI study (Chang et al., 2015). Our findings suggest that the cerebellum contributes to context-related effects of extinction. Our attempts, however, to use cerebellar transcranial direct current stimulation (tDCS) to modulate extinction and context-related extinction effects of conditioned eyeblink responses in healthy participants were largely unsuccessful (Beyer et al., 2017; Lipp et al., 2019). Lack of robustness and reproducibility of cerebellar tDCS effects are increasingly recognized (Mamlins et al., 2019) and call for further methodological refinement before more firm conclusions can be drawn in the application to patient-oriented studies.

Our most recent 7T fMRI studies show that findings related to extinction of conditioned eyeblink responses equally apply to extinction of learned fear. In healthy human participants, cerebellar cortical activations related to the acquisition of learned fear responses were reversed during extinction (Ernst et al., 2019). Furthermore, we observed activation of the posterolateral cerebellar hemisphere related to the renewal of previously extinguished conditioned fear responses in the acquisition context (Timmann, 2019). In fear conditioning paradigms, the CS–US interstimulus intervals (ISIs) typically last several seconds. Therefore, event-related designs allowed us to separate cerebellar fMRI signals related to the visual CS from signals related to the subsequent US (an aversive electric shock). We found that cerebellar activation was most pronounced in unpaired CS+ trials, that is, in trials where the US was expected but did not occur (Figure 3; Ernst et al., 2019). This activation disappeared during extinction when US omission became expected. Findings agree with the assumption that prediction error drives extinction learning (Rescorla and Wagner, 1972). Among others, reward signals may play a role. The unexpected
omission of the US is rewarding, and recent studies suggest that reward signals play a role in extinction (Kalisch et al., 2019). Furthermore, the role of the cerebellum has been shown to go beyond the processing of sensory prediction errors and to include the processing of reward prediction errors (Wagner et al., 2017). The exact nature of the observed error signal in the cerebellum needs to be elucidated in future studies.

In sum, our findings provide evidence that the cerebellum is part of the brain network subserving extinction. The cerebellum likely contributes to different aspects of extinction, and different cerebellar areas appear to be involved.

**Author contribution:** All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

**Research funding:** This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through FOR1581 projects P1, P8; DFG TI 239/10-1, 10-2 as well as by Project number 316803389 SFB 1280, projects A01, A05, A06.

**Conflict of interest statement:** The authors declare no conflicts of interest regarding this article.

---

**References**


Bionotes

Onur Güntürkün
Department of Biopsychology, Faculty for Psychology, Ruhr-University Bochum, 44780 Bochum, Germany
onur.guentuerkuen@rub.de
https://orcid.org/0000-0003-4173-5233

Onur Güntürkün is a Turkish-born Professor for Biopsychology at the Ruhr University Bochum in Germany. He is kept awake with questions like: “Why do humans and other animals have asymmetrically organized brains?” or “Can different kinds of brains produce the same cognition?”. He spent years of his life in different universities and science institutions on five continents. He uses mostly humans and pigeons as his experimental subjects but also loves to do science with dolphins and magpies. He would call himself a Cognitive and Comparative Neuroscientist who works with research approaches that can reach from simple field work, single cell recordings, and detailed neuroanatomy up to novel methods of brain imaging at ultrahigh magnetic fields.

Maik C. Stüttgen
Institute of Pathophysiology, University Medical Center of the Johannes Gutenberg University, Mainz, Germany
Maik.Stuettgen@uni-mainz.de

Maik C. Stüttgen studied psychology and neurosciences at the Universities of Giessen and Tübingen, respectively. He conducted his doctoral studies on the physiology of perception in the rat whisker system under the supervision of Cornelius Schwarz at Tübingen University, obtaining his PhD in 2007. He then joined the group of Onur Güntürkün in Bochum and investigated single-neuron activity in the nidopallium caudolaterale of operantly conditioned, freely moving pigeons. In 2013, he spent six months with Arthur Houweling at the Erasmus Medical Center in Rotterdam to work on the effects of single-cell stimulation on the local neural network in mouse cortex. Since June 2014, he is working as assistant professor at the University Medical Center in Mainz, investigating neural and psychological processes underlying adaptive behavior.
Sarah Starosta studied psychology and cognitive neuroscience at the Ruhr University Bochum and the University La Sapienza Rome. In 2015 she received her PhD from the International Graduate School of Neuroscience at the Ruhr University Bochum where she investigated the neuronal basis of extinction learning with electrophysiological, pharmacological and behavioral experiments in pigeons under the supervision of Onur Güntürkün and Maik Stüttgen. She then went abroad with a postdoctoral fellowship to work in Adam Kepecs’ group (first at the Cold Spring Harbor Laboratory, NY; now at Washington University, St. Louis, MO) where she is investigating the behavioral and neuronal algorithm that underlies decisions in a foraging setting. Her general focus of interest is how neuronal circuitry gives rise to complex phenomena, such as learning and decision making in health and disease, e.g. in the context of psychiatric diseases as well as cancer.

Roland Pusch studied biology and social sciences at the University of Bonn. In 2013, he obtained his PhD for investigating the physiology of the visual and the active electrosensory system of weakly electric fish in the department of Neuroethology – Sensory Ecology at the University of Bonn. He then joined the Biopsychology lab of Onur Güntürkün at the Ruhr University Bochum to study the neuronal fundamentals of categorization behavior and extinction learning in operantly conditioned, freely moving pigeons. In his work, he places special emphasis on sensory aspects of these complex behaviors.

Meng Gao finished her PhD study in the Biopsychology lab at the Ruhr University Bochum. During her PhD studies, she was interested in the memory mechanism of extinction learning from a comparative point of view. She used pharmacological injections and functional MRI approach to study the neural circuits of extinction learning in the pigeon brain.

Michael Nitsche is a neurologist, and psychologist, who studied at the Georg-August-University in Göttingen, and conducted his dissertation on structural connectivity of the ventral striatum at the Max Planck-Institute for Biophysical Chemistry. He worked for 16 years as a clinical neurologist, and researcher at the Department Clinical Neurophysiology of the Göttingen University Medical Center (head: Prof. Walter Paulus). Since 2015, he is Professor, and Scientific Director of the Department of Psychology and Neurosciences at the Leibniz Research Centre for Working Environment and Human Factors in Dortmund. His main research interests are the physiological foundations of cognition, and behavior in humans, with a specific dedication to learning, and memory formation. His methodological foci are non-invasive brain stimulation, pharmacological, and functional imaging approaches. For his work on non-invasive brain stimulation, he received the Alois Kornmüller, and Richard Jung awards of the German Society for Clinical Neurophysiology and Functional Imaging.

Thomas M. Ernst is an MR physicist who started out in preclinical imaging and joined Dagmar Timmann’s lab at the Neurological department in Essen in 2014. He now predominantly works at ultrahigh field strength at the Erwin L. Hahn Institute of Magnetic Resonance Imaging, where the main focus of his work has been on functional imaging of the human cerebellum during associative learning tasks and the extinction of learned behaviors.
Mark E. Ladd
Division of Medical Physics in Radiology,
German Cancer Research Center, 69120 Heidelberg, Germany
mark.ladd@dkfz-heidelberg.de

Mark E. Ladd received the B.S. degree from the University of Michigan, Ann Arbor, in 1989, the M.S. degree from Stanford University, California, in 1991, and the Ph.D. (Dr. sc. techn.) from the Swiss Federal Institute of Technology (ETH), Zurich, in 1998, all in electrical engineering. He is head of the Division of Medical Physics in Radiology at the German Cancer Research Center (DKFZ) in Heidelberg, Germany, since 2013. His research includes methodological advances in magnetic resonance imaging and spectroscopy, including imaging with ultra-high magnetic fields, parallel transmission, MRI safety, and magnetic resonance-guided radiotherapy. He is author of over 300 scientific articles and book chapters. Prof. Ladd is member of the International Society for Magnetic Resonance in Medicine (ISMRM) as well as the German Society for Medical Physics (DGMP), where he is currently president. In 2019, he was nominated for the German President’s Award for Innovation in Science and Technology (Deutscher Zukunftspreis).

Harald H. Quick
Erwin L. Hahn Institute for Magnetic Resonance Imaging, University of Duisburg-Essen, 45141 Essen, Germany
High-Field and Hybrid MR Imaging, Essen University Hospital, 45147 Essen, Germany
harald.quick@uni-due.de

Harald H. Quick has a background in Biomedical Engineering and has been appointed as Professor of High-Field and Hybrid MR. Imaging at the University Hospital Essen in 2014. He is Director of the Erwin L. Hahn Institute for MRI of the University Duisburg-Essen, a 7-Tesla ultrahigh-field MRI facility. Since 25 years he conducts research in the field of MRI. His main research foci are hardware and methods development and clinical application of (f)MRI, 7-Tesla MRI, and hybrid PET/MR. In the context of SFB 1280 his interest is to provide the best possible data available from 7T fMRI and to guarantee comparability of MRI data in multicenter fMRI studies. Before his current appointment, he was appointed full professor for MR Imaging and Deputy Director at the Institute of Medical Physics, University of Erlangen-Nürnberg. The University Hospital Essen, the Johns Hopkins University in Baltimore, and the University Hospital in Zurich were further stations of Harald Quick’s academic pathway.

Dagmar Timmann
Department of Neurology, Essen University Hospital, University of Duisburg Essen, 45147 Essen, Germany
dagmar.timmann-braun@uk-essen.de

Dagmar Timmann is a certified neurologist who works as a clinician scientist. She has been appointed as Professor of Experimental Neurology at the University Hospital of the University of Duisburg-Essen in 2000. She is head of the Ataxia Clinic in Essen for almost 20 years, and an Associate Principle Investigator at the Erwin L. Hahn Institute of Magnetic Resonance Imaging since 2016. For almost three decades, she is interested in the physiology and pathophysiology of the cerebellum. Her focus is on human cerebellar lesion studies, and structural and functional MRI of the cerebellum, studying the contribution of the cerebellum to motor performance, motor learning and cognition. She was a postdoctoral researcher in several labs in the United States and Canada, and more recently a visiting scientist at the University of Minnesota.