



Universiteit
Leiden
The Netherlands

Seeing voices: the role of multimodal cues in vocal learning

Varkevisser, J.M.

Citation

Varkevisser, J. M. (2022, October 20). *Seeing voices: the role of multimodal cues in vocal learning*. Retrieved from <https://hdl.handle.net/1887/3483920>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3483920>

Note: To cite this publication please use the final published version (if applicable).

Chapter 2

Multimodal cues in songbird vocal learning provide perspective on discrepancies between live and audio-only tutoring

Judith Varkevisser, Ralph Simon, Ezequiel Mendoza, Constance Scharff, Wouter Halfwerk & Katharina Riebel

Abstract

Bird song is a particularly well-characterised example of a socially learned vocal behaviour in non-human animals with striking analogies to human speech acquisition. Bird song learning is highly accessible to experimental manipulation, and audio playback experiments have been instrumental in the study of song learning. However, many songbird species learn less well from song playbacks than from live tutors. It is often assumed that this is because social interaction with a tutor is essential for song learning. This view has been criticised by several authors, stressing the differences between live and tape tutors in non-social dimensions such as contingencies and variability. We here want to raise awareness for the unimodal versus multimodal contrast between tape and live tutors, that constitutes for an additional overlooked dimension in this debate. Birdsong, like many animal signals and human speech, is accompanied by visual components, and thus a multi- rather than unimodal signal. A case in point is the zebra finch, *Taeniopygia guttata*, the foremost neuroethological model for vocal learning and an often-cited example for the importance of social interactions in song learning. Reviewing zebra finch song learning studies shows that research to date has not systematically differentiated between ‘social’ and ‘multimodal’ tutoring, but outcomes are often in line with the hypothesis that vocal learning may be facilitated by multimodal experiences with the signal. We conclude with an appeal and suggestions to systematically test this hypothesis regarding fundamental mechanisms in a cultural transmission process thought to be at the base of the evolution of complex communication systems.

I Introduction










Songbirds are versatile vocal learners. Bird song is a prominent example of a vocally learned signal in non-human animals (Catchpole & Slater, 1995) and often used as a model for human speech acquisition, because of the many similarities in the development of human speech and bird song (Bolhuis, Okanoya, & Scharff, 2010; Doupe & Kuhl, 1999) and the increasing interest in understanding the role of learned communication systems in the evolution of cultural transmission and cumulative culture (Whiten, 2021). Vocal learning in birds has been more extensively studied than in any other animal group given the experimental tractability of the system and early introduction of suitable learning paradigms. Thorpe’s pioneering studies in the chaffinch, *Fringilla coelebs*, (e.g. Thorpe, 1954; fully reviewed in Riebel, Lachlan, & Slater, 2015) introduced ‘tape tutoring’ (playback of pre-recorded song via loudspeakers) in combination with the analyses of sound spectrograms for the systematic study of vocal development. Tape tutoring provides excellent stimulus control and this paradigm started modern bird song research. With increasing use of this

approach it became apparent that many species learned less well from tape than from social tutors (e.g. Kroodsmma & Verner, 1978; Thielcke, 1984; Baptista & Petrinovich, 1986; Kroodsmma & Pickert, 1984; Waser & Marler, 1977), presumably because the signal lacked social salience as tutees could not socially interact with these tape tutors (Baptista & Gaunt, 1997; Catchpole & Slater, 1995; Slater, Eales & Clayton, 1988; Soma, 2011). However, other differences between live and tape tutors could be decisive and the all-importance of social interaction has remained a debated issue (e.g. Beecher, 2017; Nelson, 1997, 1998). For example, tape tutoring often consists of non-interactive exposure to looped song sequences and is thus more stereotyped and lacking the context, diurnal variability and possible contingencies of real singing (e.g. Baptista & Gaunt, 1997; Beecher, 2017; Houx & ten Cate, 1999a; Nelson, 1997, 1998). A dimension that has seen little systematic attention, is that birdsong, like many mating signals, is often multimodal (Halfwerk et al., 2019; Heberts & Papaj, 2005; Partan & Marler, 1999; Rowe, 1999) whereas a (classic) tape tutor is unimodal, providing audio-only exposure (see Table 1). We suggest a systematic investigation into the presence of multimodal cues during song exposure as an alternative (non-mutually exclusive) explanation for the often improved learning with live compared to tape tutors. Singing is accompanied by visual components, such as beak, throat, head and body movements. From other contexts, it is well documented that multi-component signals (in one or multiple modalities) can increase salience by improving detection and memorisation by receivers compared to single component signals (reviewed in Rowe, 1999). Bats, for example, learn to avoid warning signals of noxious fireflies faster with multi- than with unimodal warning signals (Leavell et al., 2018). Starlings, *Sturnus vulgaris*, perform better in temporal order judgement when auditory stimuli are preceded by visual cues (Feenders, Kato, Borzeszkowski, & Klump, 2017). Auditory filial imprinting in birds is enhanced with visual stimulation (van Kampen & Bolhuis, 1991; van Kampen & Bolhuis, 1993) and nightingales, *Luscinia megarhynchos*, learn songs from audio-only playback less well than songs combined with light flashes (Hultsch, Schleuss, & Todt, 1999).

In 6 month old human infants, fixating more on mouth movements during interactions is associated with higher levels of expressive language at age 2 (Young, Merin, Rogers, & Ozonoff, 2009) and visual speech enhances learning of phoneme contrasts (Teinonen, Aslin, Alku, & Csibra, 2008). In 12-month-olds, hearing an unknown rather than their own language increases how much infants watch a speaker's mouth (Lewkowicz & Hansen-Tift, 2012). Inspired by the human literature, and given the well-established parallels between human speech and avian song acquisition (Bolhuis et al., 2010; Doupe & Kuhl, 1999;

Soha & Peters, 2015), this review investigates whether uni- versus multimodal exposure could have been a systematic confound of comparisons of live versus tape tutors by focussing on the foremost neuroethological model for avian vocal learning, the zebra finch (Griffith & Buchanan, 2010; Mello, 2014).

Table 1. Overview of the sensory stimulation, contingencies and social interactions experienced by tutees exposed to a live tutor versus different types of tape tutoring methods. In this table, ‘sensory stimulation’ refers to auditory, visual, tactile or olfactory sensory exposure to an adult conspecific tutor, ‘contingencies’ refer to auditory, visual or tactile actions (by the tutor or the tutee) that can predict exposure to tutor song and ‘social interactions’ refer to social companionship, auditory social tutor-tutee interaction or visual social tutor-tutee interaction.

Tutoring method	Sensory stimulation				Contingencies*			Social interactions		
								company		
Live	+	+	+	+	+	+	+	+	+	+
Passive tape	+									
Operant tape	+						+			
Tape with vocal interaction	+								+	
Tape with visual stimulation	+					+				

*contingencies can arise if tutor behaviour reliably predicts song or if song is triggered by vocal, visual or physical actions by the tutee.

II Vocal learning: the zebra finch as model

Both sexes in zebra finches have an extensive call repertoire, but only males produce courtship song which consists of a string of motifs, i.e. a stereotyped sequence of individual sound elements called syllables (see Figure 1A). Motifs are learned with varying accuracy from adult conspecifics (see Figure 1B) during a sensitive period lasting roughly from 20-65 days post hatching (dph) (Eales, 1985; Gobes, Jennings, & Maeda, 2017; Immelmann, 1969; Jones, ten Cate, & Slater, 1996; Roper & Zann, 2006). Without suitable models, zebra finches develop impoverished song (reviewed in Slater et al., 1988). The exact mechanisms underlying differences in copying between individuals within and between tutoring settings are an unresolved issue (Derégnaucourt, 2011; Gobes et al., 2017; Houx, Feuth, & ten Cate, 2000; Slater, Eales, & Clayton, 1988). Next to the white crowned sparrow (Baptista & Petrinovich, 1986), zebra finches are almost generically referred to exemplify impoverished learning from tape versus live tutoring (reviewed in Derégnaucourt, 2011; Slater, Eales, & Clayton, 1988). The generally favoured explanation for this is that social interaction with the tutor facilitates song learning (e.g. Chen, Matheson, & Sakata, 2016; Derégnaucourt, Poirier, van der Kant, & van der Linden, 2013; Slater et al.,

1988, but see e.g. Nelson, 1997).

We here systematically revisit the extensive literature on zebra finch song learning (Gobes, Jennings, & Maeda, 2017; Griffith & Buchanan, 2010; Slater, Eales, & Clayton, 1988) to check whether social and multimodal tutoring were always combined or whether studies manipulated these dimensions separately. The results suggest that future studies should more systematically study whether multimodal cues enhance song learning independently or on top of social properties of a live tutor.

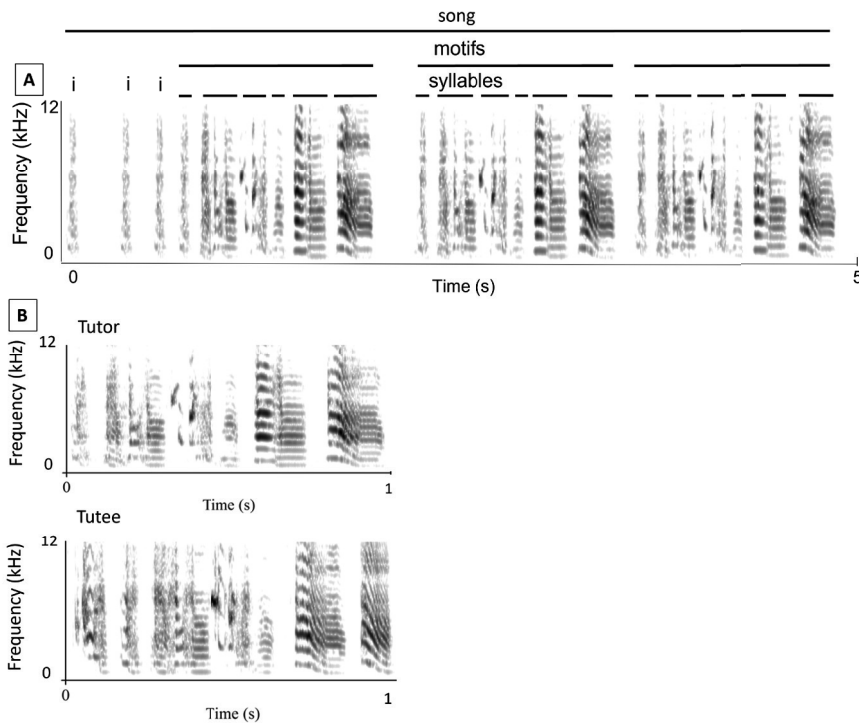


Figure 1. (A) Spectrogram of zebra finch song (also referred to as strophe) with three motif repetitions. Introductory notes are indicated with ‘i’. Song units within motifs that are separated by silent intervals are called syllables (Sossinka & Böhner, 1980). The motif of this bird consists of 6 different syllables. (B) Example of the motif of a tutor above and the motif of a tutee that was exposed to this tutor’s song during its sensitive period for song learning below. This tutee copied several syllables from the tutor. In most song learning studies, pupils received a particular type of tutoring during the sensitive period for song learning. When the pupil is adult, the amount of its song that matches the tutor song is set equal to the amount of song that the pupil has copied from the tutor. This is usually taken as a measure of song learning success (but see e.g. Geberzahn, Hultsch, and Todt (2002) that this might underestimate learning).

III The visual dimensions of singing

Acoustic signals require rhythmic mechanical movements to set the physi-

cal carrier of the signal (air, water, substrate) in motion, which often leads to visual components obligatorily coupled with specific sounds, e.g. human lip or anuran air sac movements (Bradbury & Vehrencamp, 2011). Birdsong is also accompanied by such obligatory visual components arising mainly from the beak and throat movements of song production (Goller, Mallinckrodt, & Torti, 2002; Ohms, Snelderwaard, ten Cate, & Beckers, 2010; Williams, 2001). There are additional ‘free’ signal components, like dance and wing movements, that co-occur with but are not inextricably linked to song production (Dalziell et al., 2013; Ota, Gahr, & Soma, 2015; Ullrich, Norton, & Scharff, 2016; Williams, 2001). In zebra finches, the beak and dance movements of tutors and tutees are more similar than those of unrelated males (Williams, 2001), which is consistent with tutees attending to both auditory and visual components during vocal learning. Visual cues also guide tutor choice: zebra finches preferentially choose tutors visually resembling the colour morph of birds that reared them (Mann & Slater, 1995; Mann et al., 1991). Mate recognition is also enhanced by a correct match between a male’s morph and song (Brazas & Shimizu, 2002; Campbell & Hauber, 2009). The combination of visual and auditory information can thus influence bird behaviour and it can facilitate song learning (Hultsch et al., 1999; Todt, Hultsch, & Heike, 1979). In view of these observations, we hypothesize that multimodality of a live tutor might facilitate song learning independently of the live tutor’s ‘social properties’. In the following sections, we revisit the zebra finch song learning literature comparing tutoring paradigms to ask whether studies conclusively show that social interactions - rather than differences in (multimodal) stimulus properties of tutor song - increase the salience of song models.

IV Comparing live and tape tutors across modalities

Auditory modality

Live and tape tutors could in principle provide identical auditory input, e.g. if playback is established via an audio link from a live tutor. However, most playback tutoring repeats pre-recorded song sequences (e.g. Adret, 1993a; Houx & ten Cate, 1999b; Tchernichovski, Mitra, Lints, & Nottebohm, 2001). In stark contrast, live tutors vary pitch, tempo, amplitude, number of syllables, motifs and introductory notes during singing (Glaze & Troyer, 2006; Helekar, Marsh, Viswanath, & Rosenfield, 2000) and differently so to male, female or juvenile audiences (Chen et al., 2016; Hyland Bruno & Tchernichovski, 2019; Jesse & Riebel, 2012; Williams, 2004). Variable song exposure increases stimulus engagement in females (Collins, 1999) and might prevent habituation (Krebs & Kroodsma, 1980), but we are unaware of systematic investigations whether and how variability of the stimulus affects song learning. Studies exposing tutees

to live tutors and using audio links to yoked controls (Chen et al., 2016; Eales, 1989) can provide clues here, as they eliminate differences in song variability among live and tape tutors. In these two studies, young birds learned more from live tutors (Chen et al., 2016; Eales, 1989). An observational study suggests that variability of tutor song predicts how well biological and foster offspring learn (Tchernichovski, Eisenberg-Edidin, & Jarvis, 2021). The next step will be to find out whether variability of tutor song is causal or covarying with tutor (song) properties. Future studies could test this by manipulating stereotypy and variability between and within different tutoring methods.

Visual modality

Visual access to the tutor could affect song learning via several processes. Seeing a tutor provides social stimulation and social relevance of the auditory stimulus, which might benefit the tutee's welfare or motivation to learn. Seeing the visual correlates of song production might increase the salience of the auditory signal and draw the pupil's attention towards the song or, as in humans, birds might experience improved reception and perception: in noisy conditions, seeing orofacial articulatory movements improves speech perception (Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007).

In several experiments, visual tutor models, such as pictures (Funabiki & Funabiki, 2008) or plastic models (Tchernichovski et al., 2000), were offered during passive or operant tape tutoring. However, these studies did not test whether adding visual models improved song learning and had no control groups without visuals. Deshpande and colleagues (2014) used a minimalistic single session operant tutoring design (75 s tutoring during a 2 h session using a single tutor across tutees) to capture song template formation. During a single session (at 35 or 45 dph), tutees had operant control over audio or audio-visual song playbacks where the visual stimulus was either preceding (VA), concurrent with or following audio (AV) playback. Only tutees from the simultaneous audio-visual or AV conditions showed significantly more song learning than untutored birds. Song learning was low overall, probably due to the minimal exposure level. However, the results support the hypothesis that visual exposure to a singing tutor can facilitate song learning. Further studies using replicate high-quality video tutors throughout the sensitive period could systematically test whether the addition of a video to a tape tutor leads to increased song learning outside a single session tutoring design. Price (1979) compared song learning in five males raised with two live tutors behind a visual separation to learning in a male that was not visually separated from a live tutor. The visually separated birds copied few song elements, while the male with visual access

produced only elements copied from its tutor. Although it is hard to draw conclusions from such a small sample size, this study does suggest that visual exposure to a tutor facilitates song learning in zebra finches.

Other modalities

Multicomponent signals generally lead to better detection and learning than single component signals (reviewed in Rowe, 1999). Olfactory and gustatory cues enhance the learning of visual warning signals in chicken (Rowe & Guilford 1996). Zebra finches show olfactory guided natal nest and kin recognition (reviewed in Krause et al., 2018). Hence, stimulation in other than auditory and visual modalities might influence song learning. It is feasible that olfactory components of the tutor guide tutor choice, or reinforce proximity, but as yet this has not been studied. Studies have, however, investigated the effect of physical interactions on song learning. Eales (1985) suggested that physical contact with a live tutor is unnecessary for song learning after observing song learning in tutees separated from their tutor by wire mesh. However, Adret (1992) found poor copying from a tutor at 50 cm away and hypothesized that this was due to a lack of physical interaction. We could not find studies directly comparing song learning in tutees that could or could not physically interact with the same tutor. This issue is far from resolved, as physical tutor-tutee interactions are frequent (Adret, 2004; Clayton, 1987; Mann & Slater, 1995; Morris, 1954). Future studies will thus have to investigate a potential influence of olfactory and tactile cues on song learning.

V Contingencies with tutor song exposure: live versus tape tutors

During live tutoring, the tutor's or tutee's behaviour might predict tutor song exposure, possibly facilitating attention to tutor song and song learning. For instance, tutor song might always be preceded by specific tutor behaviour. The absence of such contingencies is a pronounced and systematic difference between tape and live tutoring (Houx & ten Cate, 1999a; ten Cate, 1991).

Vocal contingencies with tutor song

Searching for contingencies between tutor behaviour and singing, Houx and ten Cate (1998) found the only tutor behaviour to predict singing was tutor song itself, as tutors usually produce song in bouts. Calls were not recorded in this study, but might also provide vocal contingencies with tutor song. Tutees that had an audio-link to a live tutor were passively exposed to all of the tutor's vocalizations, but still copied less from the tutor than tutees co-housed with the tutor (Chen et al., 2016; Eales, 1989). Evidence from other species suggests such contingencies might affect song learning – white crowned sparrows learn

better if songs start with a specific acoustic cue (Soha & Marler, 2000).

Visual contingencies with tutor song

Investigating the effect of visual contingencies with tutor song on zebra finch song learning, two studies found that groups receiving visual stimulation contingent on song did not learn more than those that received audio playback only (Bolhuis, van Mil, & Houx, 1999; Houx & ten Cate, 1999a): the visual stimulus was a taxidermic mount of an adult male zebra finch that was revealed right before or after song playbacks (Houx & ten Cate, 1999a) or in another study during and after playbacks (Bolhuis, van Mil, & Houx, 1999). Although a cylinder that was raised and lowered provided some motion, the taxidermic mount itself was stationary (Bolhuis et al., 1999; Houx & ten Cate, 1999a). A temporary coupling between auditory and visual stimulation may be necessary to facilitate song learning. Nightingales (Hultsch et al., 1999), for example, learned better from song playbacks presented with a synchronously flashing stroboscope than from songs presented without stroboscope flashes. A similar temporary coupling of auditory and visual stimulation might occur in the beak or body movements of a singing bird. We thus hypothesize that visual contingencies with song exposure might be one of the factors facilitating song learning from a live tutor.

Contingencies with tutee behaviour

Bird song and language learning have both been suggested to be a form of operant conditioning, where young learners experience reinforcement when motivated to hear vocalizations and actively elicit exposure to speech or song (language: Sturdy & Nicoladis, 2017, birdsong: Adret, 1993a). Operant tasks allow to test these ideas experimentally. The most powerful tests to date used paired designs, where one bird could actively trigger playback by perch hopping or key pecking and a yoked control simultaneously received the same playback via a second loudspeaker. Experimental birds (with operant control over playback) copied more song than their yoked controls in one study (Adret, 1993), but not in other studies (Houx & ten Cate, 1999b; ten Cate, 1991), possibly because otherwise single housed experimental and control birds could vocally interact in one study (Adret 1993). The interaction between the birds in combination with exposure to the operant tape tutor, might have affected song copying success (Houx & ten Cate, 1999b). However, small sample sizes ($n = 3$ in Adret (1993)), and lack of a live control condition make it difficult to settle these questions. Several subsequent studies have successfully used operant tutoring and reported substantial song imitation compared to untutored birds (Derégnaucourt, Mitra, Fehér, Pytte, & Tchernichovski, 2005; Phan, Pytte, &

Vicario, 2006; Tchernichovski, Mitra, Lints, & Nottebohm, 2001). Unfortunately, none of these studies included a control treatment involving passive audio exposure to the songs used for the operant tutoring, which supposedly leads to higher tutor song similarity than in untutored birds (Chen et al., 2016). Additional differences also hamper direct comparison between these operant and standard tape tutoring paradigms. First, there was potential additional visual stimulation, as the playback loudspeaker was hidden inside a plastic zebra finch model in the tutee's cage. Second, restricted song exposure (maximum 20 reinforced key pecks per day) creates a variable reward schedule which can be more reinforcing than continuous rewarding (Ferster & Skinner, 1957).

We found one study directly comparing live, operant and tape tutors, where the tutees in the operant and passive tape tutoring condition also got restricted song exposure (Derégnaucourt et al., 2013). Birds from these two conditions, once adult, had a lower similarity to the tutor song than birds from the live condition. However, there were several differences between the live and the operant or passive tape tutoring paradigm (for instance, variable versus stereotyped song exposure and raised with a social companion versus in social isolation, see Table 2), making it difficult to discern which factors contributed to this difference. The songs of the operantly trained birds were significantly more similar to the tutor song than the songs from the passively exposed birds. This suggests that active control over song exposure or the partial reward scheme positively affected song learning.

VI Social companionship versus social isolation

An aspect of live tutoring that has to date seen little attention is that the mere presence of a companion could affect song development. Live tutors provide social company while tape-tutored birds are often housed in social isolation, which could affect a bird's hormonal and physiological state and consequently song learning. Adret (2004), for example, observed that tutees co-housed with a female in addition to their tutor learned better than those without. Also, birds reared in song isolation show better song learning when reared with peers than when reared alone (Jones et al., 1996). While this is generally interpreted as using other isolate males' song as model, the presence of social companions could improve song learning by several other mechanisms. In social animals like zebra finches, social isolation might have a negative effect on welfare (e.g. corticosterone levels in the blood of zebra finches that were socially isolated for 10 minutes are significantly higher than baseline levels (Banerjee & Adkins-Regan, 2011)) and on the motivation to practice and learn song. Zebra finches produce more song with a male or female companion, compared to socially

isolated housing (Jesse & Riebel, 2012). Less singing could reduce practicing during motor learning. In young zebra finches that produce immature songs, a female conspecific can elicit songs with more mature properties (Kojima & Doupe, 2011). Companions could thus lead to more practice which is crucial as demonstrated by temporarily pharmacological blocking of vocal motor control during late motor practice (but not other ages) which impairs learning (Pytte & Suthers, 2000). Social companions could not only encourage practice but also guide song development. In cowbirds, *Molothrus ater ater*, non-singing females shape male song production (West & King, 1988) and there is increasing evidence that female zebra finches might affect male song learning (Carouso-Peck & Goldstein, 2019; Jones & Slater, 1993; Williams, 2004). For example, zebra finches were found to learn better if housed with a companion than when housed alone or with a deaf female companion (Williams, 2004). Also, non-vocal feedback (fluff-ups performed by the mother before, during or after tutee song production) was positively correlated with song learning success (Carouso-Peck, Menyhart, DeVogd, & Goldstein, 2020). This suggests that reactions of companions to the tutee's song play a role in the song learning process.

Rearing in isolation from conspecifics also affects adult auditory discrimination, e.g. birds reared in isolation perform worse in auditory discrimination tasks than socially reared birds (Sturdy, Phillmore, Sartor, & Weisman, 2001). To investigate the effect of social interaction with a tutor on neuronal responsiveness in the auditory cortex, juvenile zebra finches were exposed to playbacks of their tutor's song while in social isolation or paired with their tutors (Yanagihara & Yazaki-Sugiyama, 2019). In the juveniles paired with their tutor, but not in the juveniles in social isolation, neurons exhibited selective auditory responses to the playbacks. Social isolation was only compared with the tutor-present situation, making it impossible to disambiguate whether this effect arose from having a social companion in general, or more specifically from being able to socially interact with the tutor during song exposure. Albeit as yet not a subject of systematic study, the combined indirect evidence from the studies discussed above suggests that social companionship per se, which is absent in standard tape tutoring settings, may affect song development and learning in zebra finches. Future studies should aim for a comparable social environment in different tutoring treatments.

VII Tutor reaction to tutee song

In live tutoring paradigms, tutors might respond vocally or visually to tutee vocalizations, thereby reinforcing particular song elements or singing behaviour. Tape-tutored tutees do not receive tutor feedback on their vocalizations.

Tutor reacts vocally to tutee song

Among adult zebra finches, vocal interactions can influence singing: males are more likely to alter or end a motif if a female conspecific calls while they are singing (Williams, 2004). While the idea of reinforcement by vocal tutor reactions is appealing theoretically, detailed observations of tutor/tutee interactions showed no vocal contingent tutor reactions, defined as tutor behaviour occurring more often within 15 seconds after tutee song onset than expected by chance (Houx & ten Cate, 1998). However, the magnitude of father's singing responses to their sons singing was positively correlated with sons' song learning accuracy in another study (Carouso-Peck et al., 2020).

Eales (1989) compared song learning in zebra finches in three different treatments: birds that could visually and vocally interact with a tutor in an adjoining cage all learned at least some elements from this tutor. In a group that could interact only vocally with a tutor behind an opaque screen, four birds copied elements from the tutor, while three birds copied elements they heard before 35 dph. None of the birds that could only hear the tutor's song from a loudspeaker copied from it. This suggests that vocal interaction facilitated song learning to some degree. However, as pointed out by Nelson (1997), interpretation of these results is difficult as the birds that could vocally and visually interact with the tutor were housed in a room with many conspecifics, while birds in the other two groups were housed in sound-isolation boxes. Using both a one-way and two-way audio-link, i.e. one link that gave a tutee passive tutor exposure and one that allowed vocal tutor-tutee interaction, Chen et al., (2016) compared song learning between vocally interactive and non-interactively tutored birds, but found no differences in song learning success. Song copying was, however, poor in all groups, as exposure was limited to only one day. These results are therefore not comparable to standard live tutoring situations. Further studies are necessary to find out to what extent vocal exchanges between tutors and tutees affect vocal learning.

Tutor reacts visually to tutee song

An observational study found no tutor behaviour to occur more often than expected after tutee song (Houx & ten Cate, 1998). There is no experimental study investigating this question involving male tutors, but female cowbirds reinforce the production of specific song elements by wing movements (West & King, 1988). Likewise, young zebra finches receiving contingent visual feedback (video playback) of fluff-up behaviour by a female on their immature song production (Carouso-Peck & Goldstein, 2019) were found to copy more tutor

song than birds receiving non-contingent visual feedback. Live tutors could also provide visual feedback to song learning juveniles. This might be one of the factors facilitating song learning from a live tutor.

VIII Is social interaction or multimodal exposure the key difference between live and tape tutors?

As evidenced from the previous paragraphs and Table 2, we have as yet not unambiguously identified which stimulus properties of a live tutor improve learning compared to tape tutoring paradigms. Owing to the logistics of tutoring experiments and research interests, most studies investigated single stimulus dimensions, but the facilitating effect of a live tutor might arise from a combination of factors. The most favoured hypothesis in the literature is that the social interaction between tutor and tutee is decisive for song learning (Baptista & Petrinovich, 1986; Catchpole & Slater, 1995; Slater, Eales, & Clayton, 1988) and also for other forms of channelled learning such as filial and sexual imprinting (ten Cate, 1994). A mechanism here might be that social interactions enhance attention (Chen et al., 2016) or engage birds in more practice (Jesse & Riebel, 2012) and thereby promote learning. This raises the question which stimulus properties of a ‘social interaction’ are decisive (Nelson, 1997). Table 2 shows that many studies did not systematically control for the fact that live song exposure is multimodal whereas taped song is not. Much of the evidence for ‘social interaction’ with careful re-evaluation could also be interpreted as evidence for higher stimulus salience arising from multimodality or contingencies. Moreover, more time with a stimulus leads to more interactions – we can thus not establish any causality from observations showing that birds learned more from the male they interacted with most (Eales, 1987; Williams, 1990) or that showed most aggressive behaviour towards them (Clayton, 1987, but see Houx & ten Cate, 1998; Mann & Slater, 1995; Williams, 1990), as this could be a matter of total time in close proximity (Slater & Richards, 1990; Mann & Slater, 1995). Active approaches between tutors and tutees after the tutor started singing (Houx & ten Cate, 1998) can be interpreted as social attraction, but also as attraction to the multidimensional properties of song. Closer and longer proximity and more interaction also mean more opportunity for the tutee to see the tutor singing and these observations thus support either the social or multimodal hypothesis. Similarly, the activation of a mesocortical dopamine circuit by the presence of a singing tutor might be related to social aspects of the tutor but also to multimodal exposure to it (Tanaka, Sun, Li, & Mooney, 2018).

Some studies have tried to address the effect of social interaction with a tu-

tor on zebra finch song learning: testing whether social interaction affects the duration of the sensitive period for song learning, Morrison and Nottebohm (1993) compared tutees visually separated from their tutor by cardboard dividers, social + song isolated tutees and aviary-reared control birds. Unlike the control birds, the other two groups learned from a new unrelated tutor beyond the normal closure of the sensitive phase, which the authors interpreted to show that lack of visually guided social interactions delayed the closure of the sensitive phase. However, the mere visual exposure to the singing tutor could have been crucial. Chen et al. (2016) housed young zebra finches in social isolation during the sensitive period, except for five days when a live tutor was housed next to them. Individuals in one treatment could vocally and visually interact with the tutor, while individuals in a yoked control treatment could only hear the tutor from a loudspeaker. Song in these birds was more similar to the tutor song than song in untutored birds, and more so in the live than in the passively tutored birds. Chen et al. (2016) conclude that higher song copying success in the live tutored group is caused by the social interactions with the tutor in this condition, but again, the results are also in line with the hypothesis that song learning could improve because of multimodal exposure to the tutor. A study investigating song learning in zebra finch tutees housed in an aviary with peers, with a visible and audible adult tutor housed in a separate cage outside of the aviary, found that the adult song of the tutees resembled the song of their peers more than the song of their tutor (Honarmand, Riebel, & Naguib, 2015). This suggests that the ability to socially interact with peers made the tutees more likely to learn from them than from the tutor. We agree that social interaction is probably important for song learning, however, we argue that simpler mechanisms such as visual exposure to a tutor should be investigated as possible additional contributing factors to zebra finch song learning.

IX A proposed framework for disentangling multimodal and social dimensions

Overall, studies found more tutor song copying by live than tape-tutored tutees (see Table 2) with the exception of one study using a plastic tutor bird and restricted song exposure (Phan et al., 2006). From the many studies and different learning outcomes, we could not identify one single key factor systematically associated with the difference in song learning from live or tape tutors. Probably, a combination of factors associated with a live tutor has a facilitating effect on song learning. As we hypothesized, one of these factors might be that a live tutor offers multimodal exposure to song. Revisiting the literature has shown great asymmetries in uni- versus multimodal tutoring regimes for the majority of non-live tutoring approaches (see Table 2). Song tutoring studies have not

systematically controlled for this dimension and the results of many studies are in line with both the interpretations that social interaction or multimodal exposure to song facilitated song learning. Several authors have pointed out the potential importance of both auditory and visual stimulation for zebra finch song learning (Adret, 1992, 1997; Bolhuis et al., 1999; Carouso-Peck & Goldstein, 2019; Houx & ten Cate, 1999a). However, as Table 2 shows, in all studies to date (but Deshpande et al. 2014) visual interaction with a tutor and multimodal exposure to tutor song were confounded: in all conditions with auditory and visual tutor exposure, tutees were also able to socially interact with the tutor. Future studies should thus aim at investigating visual interaction and multimodal exposure separately to test their effects on song learning.

To investigate potential beneficial effects of multimodal tutor exposure on song learning, live tutored birds should be compared with birds that can also see and hear a singing tutor, but without being able to socially interact with it. This can be achieved with one-way audio-visual links, for example using a one-way mirror such that a tutee can see the tutor, while the tutor cannot see the tutee. If this tutee copies more from the tutor than a tutee receiving the same auditory input without visual access to the tutor, this might indicate that multimodal cues facilitate song learning. However, it is then impossible to disentangle the effect of mere visual exposure to a moving and singing adult, which might affect the tutee's general motivation or attention, from the effect of exposure to the specific visual correlates of song production, e.g. the tutor's beak and head movements. To focus on exposure to the visual component of song, one could use artificial tutors, such as videos of singing zebra finches. Creating these videos is relatively easy, although one should be aware that standard video systems are developed for human vision, which makes it important to adjust e.g. colours and frame rate to make them suitable for bird vision (Chouinard-Thuly et al., 2017; Fleishman & Endler, 2000; Oliveira et al., 2000; Tedore & Johnsen, 2017). We are aware of three studies that have tutored zebra finches with videos (Adret, 1997; Deshpande, Pirlepsov, & Lints, 2014; Ljubičić, Hyland Bruno, & Tchernichovski, 2016). Passive and operant exposure to a video tutor led to impoverished adult songs in the pioneering study by Adret (1997), who worried that the sound playback through the TV monitor loudspeakers was distorted and so impaired learning. With very low amounts of exposure to another video tutor (only 75 seconds in total), there was very little learning (Deshpande et al., 2014). Finally, preliminary results suggest that zebra finches can adapt the pitch of already learned syllables according to the song of a video tutor singing toward the tutee (Ljubičić et al., 2016). However, as yet no study has compared birds exposed to the same live and videotaped tutor.

An important caveat regarding videos studies is their limited two-dimensional representation of a bird. The lack of depth might influence the salience of visual cues and thereby their effect on song learning. This could be overcome by using moving 3-dimensional models, such as robotic birds which have already been used successfully in studies on the importance of multimodal song in territory defence (Anderson, DuBois, Piech, Searcy, & Nowicki, 2013, Ręk & Magrath, 2016), suggesting they could be applied in song tutoring studies as well (Simon et al., 2019).

This review mainly focused on production learning, but additional perspectives arise from preference learning. Both male and female zebra finches develop preferences for tutor over unfamiliar songs from live (Riebel, Smallegange, Terpstra, & Bolhuis, 2002) and tape tutors (Holveck & Riebel, 2014; Houx & ten Cate, 1999a, b; Riebel, 2000). In tape-tutored males, song production learning (amount of elements copied) does not predict preference strength for this tutor song (Houx & ten Cate, 1999b), indicating that different mechanisms might be involved in preference and production learning. In males, no direct comparison has been made so far between preference learning from live and tape tutors. In females, preference learning did not differ between live and tape tutors (Holveck & Riebel, 2014). However, the females were housed in pairs and exposed to a series of a live tutor before 35 days and then two tape tutors between 35 and 65 days. It would be premature to conclude that preference and production learning or female and male learning differ in how susceptible they are to uni- versus multimodal – or interactive tutoring (Riebel, Odom, Langmore, & Hall, 2019). Controlled tutoring studies manipulating one modality at a time, as described earlier in this paragraph, and assessing production and preference learning can shed light on whether visual cues have an effect on the song learning process in general, and whether there is a difference in the effect of visual cues on song production and preference learning.

Table 2. Zebra finch tutoring studies comparing song copying across tutoring regimes. ‘+’ means present, ‘-’ means absent. * is a value taken from a graph, ** is calculated based on data in the article.

Multimodal cues provide perspective on live vs audio-only tutoring discrepancy

Study (in order of publication year)	Treatment groups	N	Learning success across groups ¹	Sensory stimulation				Contingencies				Social interaction				Tutoring [days]	Scoring song copying ²	Mean % ± SE pupil song copied from tutor	Mean %± SE of tutor song copied by pupil ³
				Auditory	Visual	Tactile	Olfactory	Vocal	Visual	Other	Companionship	Vocal	Visual	General					
Price 1979	1. Non-visible live tutor 2. Live tutor	5		+	-	+	+	+	+	+	+	+	+	+	+	V			
		1		+	-	-	+	+	+	+	+	+	+	+	+	+	V		
Eales 1989	1. One-way audio-link live tutor 2. Non-visible tutor 3. Live tutor	8	3 > 1+2	+	-	-	-	+	-	-	-	-	-	-	-	V			
		7	1=2	+	-	-	-	+	-	-	-	-	-	-	-	V			
		11		+	+	-	+	+	+	+	+	+	+	+	+	V			
Adret 1993	1. Tape tutor with vocal interaction with peers 2. Operant tape tutor 3. Operant tape tutor with vocal interaction with peers	3		+	-	-	-	-	-	-	-	-	-	-	-	V		39±13	
		3	2,3>1 ⁴	+	-	-	-	-	-	-	-	-	-	-	-	V		±54**	
		3		+	-	-	-	-	-	-	-	-	-	-	-	V		76±3	
Bolhuis et al. 1999	1. Tape tutor sim. with visual stimulus 2. Tape tutor followed by visual stimulus	7	1=2	+	-	-	-	-	-	+	-	-	-	-	-	V		32±11*	
		8		+	-	-	-	-	-	-	-	-	-	-	-	V		30±9*	
Houx & ten Cate 1999a	1. Tape tutor 2. Tape tutor preceded by visual stimulus 3. Tape tutor followed by visual stimulus	8		+	-	-	-	-	-	-	-	-	-	-	-	V		36±9*	
		8	1=2=3	+	-	-	-	-	-	+	-	-	-	-	-	V		25±8*	
		8		+	-	-	-	-	-	-	-	-	-	-	-	V		33±7*	
Houx & ten Cate 1999b	1. Tape tutor 2. Operant tape tutor	8	1=2	+	-	-	-	-	-	-	-	-	-	-	-	V		43±8	
		8		+	-	-	-	-	-	-	-	-	-	-	-	V		50±9*	

Adret 2004	1. Live tutor, tutee's eyes occluded	8																						42±6	
	2. Live tutor, tutee's eyes occluded, female sibling in cage	8	2>1																					63±9	
Phan et al. 2006	1. Operant tutor + plastic model	9																						61±6	
	2. Live tutor	11	1=2																					71±4	
Derégnaucourt et al. 2013	1. Tape tutor	19																						60±4*	
	2. Operant tape tutor	33	3>2>1																					61±4*	
	3. Live tutor	10																						76±4*	
Desphande et al. 2014	1. Untutored isolates	10																						13 ^{*5}	
	2A. Operant tape tutor ⁶	6																						16*	
	3A. Operant video tutor with audio and video simultaneous	6	4A>1																					16*	
	4A. Operant video tutor with audio before video	5																						20*	
	5A. Operant video tutor with video before audio	5																						17*	
	2B. Operant tape tutor	5																						20*	
	3B. Operant video tutor with audio and video simultaneous	6																						21*	
	4B. Operant video tutor with audio before video	7	3B, 4B>1																						21*
	5B. Operant video tutor with video before audio	6																							19*

Chen et al. 2016	1. One-way audio-link to live tutor	9		+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5	S	34±6*
	2. Live tutor	10		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	5	S	59±6*
	3. One-way audio-link to live tutor	6	2>1 3=4=5	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	<1	S	33±4*
	4. Two-way audio-link to live tutor	5		+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	<1	S	34±4*
	5. Live tutor	6		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	<1	S	43±4*
Carouso-Peck & Goldstein 2019	1. Live tutor (0-35 dph) and video non-contingent to tutee song	9		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	35	S	35±17
	2. Live tutor (0-35 dph) and video contingent to tutee song ⁶	9	2>1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	35	S	62±12

¹ > indicates a significant difference between groups, while = indicates that groups did not differ significantly from each other.

² V = Visual inspection of spectrogram, S = similarity score in Sound Analysis Pro

³ If original papers gave decimals, these have been rounded off to standardize display

⁴ Difference between 2 and 3 was not tested

⁵ This study only reports a similarity score where three different SAP measures (similarity, accuracy and sequential match) are combined

⁶ 2A, 3A, 4A and 5A were tutored at 35 days post-hatching, while 2B, 3B, 4B and 5B were tutored at 45 days post-hatching

X Conclusion

Re-evaluating previous song tutoring experiments showed that research to date has not systematically controlled for uni- versus multimodality of exposure when testing the importance of social interaction for zebra finch song learning. Investigating multimodal tutoring separately from social components might help to better understand the differences in stimulus properties that arise from live versus tape tutoring. By separately investigating the effect of visual cues and interaction on song learning, the contribution of multimodal cues and social interaction to the song learning process can be disentangled. Additionally, by standardizing the social environment of birds in different tutoring conditions, the effect of a social companion versus a social tutor can be investigated. Identification of relevant stimulus properties should improve our insights into the mechanisms underlying social vocal learning processes that are at the heart of the evolution of cultural transmission and cumulative culture in communication.

Acknowledgements

We would like to thank Carel ten Cate for comments on an earlier version of this manuscript and members of the Behavioural Biology group in Leiden for discussion. Funding for this review was provided by the Human Frontier Science Program (No RGP0046/2016).

References

- Adret, P. (1992). Imitation du chant chez les diamants mandarins: voir, entendre et interagir. *Annales de La Fondation Fyssen*, 7, 73–82.
- Adret, P. (1993). Operant conditioning, song learning and imprinting to taped song in the zebra finch. *Animal Behaviour*, 46, 149–159.
- Adret, P. (1997). Discrimination of video images by zebra finches (*Taeniopygia guttata*): direct evidence from song performance. *Journal of Comparative Psychology*, 111(2), 115–125. <https://doi.org/10.1037/0735-7036.111.2.115>
- Adret, P. (2004). Vocal imitation in blindfolded zebra finches (*Taeniopygia guttata*) is facilitated in the presence of a non-singing conspecific female. *Journal of Ethology*, 22(1), 29–35. <https://doi.org/10.1007/s10164-003-0094-y>
- Anderson, R. C., DuBois, A. L., Piech, D. K., Searcy, W. A., & Nowicki, S. (2013). Male response to an aggressive visual signal, the wing wave display, in swamp sparrows. *Behavioral Ecology and Sociobiology*, 67(4), 593–600. <https://doi.org/10.1007/s00265-013-1478-9>
- Banerjee, S. B., & Adkins-Regan, E. (2011). Effect of isolation and conspecific presence in a novel environment on corticosterone concentrations in a social avian species, the zebra finch (*Taeniopygia guttata*). *Hormones and Behavior*, 60(3), 233–238. <https://doi.org/10.1016/j.yhbeh.2011.05.011>
- Baptista, L. F., & Gaunt, S. L. L. (1997). Social interaction and vocal development in

- birds. In C. T. Snowdon & M. Hausberger (Eds.), *Social influences on vocal development* (pp. 23–40). Cambridge, Cambridge University Press.
- Baptista, L. F., & Petrinovich, L. (1986). Song development in the white-crowned sparrow: social factors and sex differences. *Animal Behaviour*, 34(5), 1359–1371. [https://doi.org/10.1016/S0003-3472\(86\)80207-X](https://doi.org/10.1016/S0003-3472(86)80207-X)
- Beecher, M. D. (2017). Birdsong learning as a social process. *Animal Behaviour*, 124, 233–246. <https://doi.org/10.1016/j.anbehav.2016.09.001>
- Bolhuis, J. J., Okanoya, K., & Scharff, C. (2010). Twitter evolution: Converging mechanisms in birdsong and human speech. *Nature Reviews Neuroscience*, 11(11), 747–759. <https://doi.org/10.1038/nrn2931>
- Bolhuis, J., van Mil, D., & Houx, B. (1999). Song learning with audiovisual compound stimuli in zebra finches. *Animal Behaviour*, 58, 1285–1292. <https://doi.org/10.1006/anbe.1999.1266>
- Bradbury, J. W., & Vehrencamp, S. L. (2011). *Principles of animal communication*. Sinauer Associates, Sunderland.
- Campbell, D. L. M., & Hauber, M. E. (2009). The disassociation of visual and acoustic conspecific cues decreases discrimination by female zebra finches (*Taeniopygia guttata*). *Journal of Comparative Psychology*, 123(3), 310–315. <https://doi.org/10.1037/a0015837>
- Carouso-Peck, S., & Goldstein, M. H. (2019). Female social feedback reveals non-imitative mechanisms of vocal learning in zebra finches. *Current Biology*, 29, 631–636. <https://doi.org/10.1016/j.cub.2018.12.026>
- Carouso-Peck, S., Menyhart, O., DeVogd, T. J., & Goldstein, M. H. (2020). Contingent parental responses are naturally associated with zebra finch song learning. *Animal Behaviour*, 165, 123–132. <https://doi.org/10.1016/j.anbehav.2020.04.019>
- Catchpole, C. K., & Slater, P. J. B. (1995). How song develops. In C. K. Catchpole & P. J. B. Slater (Eds.), *Bird Song: Biological Themes and Variations* (pp. 45–69). Cambridge: Cambridge University Press.
- Chen, Y., Matheson, L. E., & Sakata, J. T. (2016). Mechanisms underlying the social enhancement of vocal learning in songbirds. *Proceedings of the National Academy of Sciences*, 201522306. <https://doi.org/10.1073/pnas.1522306113>
- Chouinard-Thuly, L., Gierszewski, S., Rosenthal, G. G., Reader, S. M., Rieucan, G., Woo, K. L., & Witte, K. (2017). Technical and conceptual considerations for using animated stimuli in studies of animal behavior. *Current Zoology*, 63, 5–19. <https://doi.org/10.1093/cz/zow104>
- Clayton, N. S. (1987). Song tutor choice in zebra finches. *Animal Behaviour*, 35, 714–721. [https://doi.org/10.1016/0003-3472\(95\)80212-6](https://doi.org/10.1016/0003-3472(95)80212-6)
- Collins, S. A. (1999). Is female preference for male repertoires due to sensory bias? *Proceedings of the Royal Society B: Biological Sciences*, 266(1435), 2309–2314. <https://doi.org/10.1098/rspb.1999.0924>
- Dalziell, A. H., Peters, R. A., Cockburn, A., Dorland, A. D., Maisey, A. C., & Magrath, R. D. (2013). Dance choreography is coordinated with song repertoire in a complex avian display. *Current Biology*, 23(12), 1132–1135.

- <https://doi.org/10.1016/j.cub.2013.05.018>
- Derégnaucourt, S. (2011). Birdsong learning in the laboratory, with especial reference to the song of the zebra finch (*Taeniopygia guttata*). *Interaction Studies*, 12, 324–350. <https://doi.org/10.1075/is.12.2.07der>
- Derégnaucourt, S., Mitra, P. P., Fehér, O., Pytte, C., & Tchernichovski, O. (2005). How sleep affects the developmental learning of bird song. *Nature*, 433(7027), 710–716. <https://doi.org/10.1038/nature03275>
- Derégnaucourt, S., Poirier, C., van der Kant, A., & van der Linden, A. (2013). Comparisons of different methods to train a young zebra finch (*Taeniopygia guttata*) to learn a song. *Journal of Physiology*, 107, 210–218. <https://doi.org/10.1016/j.jphysparis.2012.08.003>
- Deshpande, M., Pirlepsov, F., & Lints, T. (2014). Rapid encoding of an internal model for imitative learning. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 281(1781), 20132630. <https://doi.org/10.1098/rspb.2013.2630>
- Doupe, A. J., & Kuhl, P. K. (1999). Bird song and human speech: common themes and mechanisms. *Annu. Rev. Neurosci.*, 22, 567–631. <https://doi.org/10.1146/annurev.neuro.22.1.567>
- Eales, L. A. (1985). Song learning in zebra finches: some effects of song model availability on what is learnt and when. *Animal Behaviour*, 33(4), 1293–1300. [https://doi.org/10.1016/S0003-3472\(85\)80189-5](https://doi.org/10.1016/S0003-3472(85)80189-5)
- Eales, L. A. (1987). Do zebra finch males that have been raised by another species still tend to select a conspecific song tutor? *Animal Behaviour*, 35(5), 1347–1355. [https://doi.org/10.1016/S0003-3472\(87\)80007-6](https://doi.org/10.1016/S0003-3472(87)80007-6)
- Eales, L. A. (1989). The influences of visual and vocal interaction on song learning in zebra finches. *Animal Behaviour*, 37, 507–508. [https://doi.org/10.1016/0003-3472\(89\)90097-3](https://doi.org/10.1016/0003-3472(89)90097-3)
- Feenders, G., Kato, Y., Borzeszkowski, K. M., & Klump, G. M. (2017). Temporal ventriloquism effect in european starlings: evidence for two parallel processing pathways. *Behavioral Neuroscience*, 131(4), 337–347. <https://doi.org/10.1037/bne0000200>
- Ferster, C. B., & Skinner, B. F. (1957). *Schedules of reinforcement*. Appleton-Century-Crofts.
- Fleishman, L. J., & Endler, J. A. (2000). Some comments on visual perception and the use of video playback in animal behavior studies. *Acta Ethologica*, 3(1), 15–27. <https://doi.org/10.1007/s102110000025>
- Franz, M., & Goller, F. (2002). Respiratory units of motor production and song imitation in the zebra finch. *Journal of Neurobiology*, 51(2), 129–141. <https://doi.org/10.1002/neu.10043>
- Funabiki, Y., & Funabiki, K. (2008). Song retuning with tutor model by adult zebra finches. *Developmental Neurobiology*, 68(5), 645–655. <https://doi.org/10.1002/dneu.20597>
- Geberzahn, N., Hultsch, H., & Todt, D. (2002). Latent song type memories are accessible through auditory stimulation in a hand-reared songbird. *Animal*

- Behaviour, 64(5), 783–790. <https://doi.org/10.1006/anbe.2002.3099>
- Glaze, C. M., & Troyer, T. W. (2006). Temporal Structure in Zebra Finch Song: Implications for Motor Coding, 26(3), 991–1005. <https://doi.org/10.1523/JNEUROSCI.3387-05.2006>
- Gobes, S. M. H., Jennings, R. B., & Maeda, R. K. (2017). The sensitive period for auditory-vocal learning in the zebra finch: consequences of limited-model availability and multiple-tutor paradigms on song imitation. *Behavioural Processes*, 163, 5–12. <https://doi.org/10.1016/j.beproc.2017.07.007>
- Griffith, S. C., & Buchanan, K. L. (2010). The zebra finch : the ultimate Australian supermodel. *Emu*, 110, v–xii. https://doi.org/10.1071/MUv110n3_ED
- Halfwerk, W., Varkevisser, J., Simon, R., Mendoza, E., Scharff, C., & Riebel, K. (2019). Toward testing for multimodal perception of mating signals. *Frontiers in Ecology and Evolution*, 7, 2013–2019. <https://doi.org/10.3389/fevo.2019.00124>
- Hebets, E. A., & Papaj, D. R. (2005). Complex signal function: Developing a framework of testable hypotheses. *Behavioral Ecology and Sociobiology*, 57(3), 197–214. <https://doi.org/10.1007/s00265-004-0865-7>
- Helekar, S. A., Marsh, S., Viswanath, N. S., & Rosenfield, D. B. (2000). Acoustic pattern variations in the female-directed birdsongs of a colony of laboratory-bred zebra finches. *Behavioural Processes*, 49(2), 99–110. [https://doi.org/10.1016/S0376-6357\(00\)00081-4](https://doi.org/10.1016/S0376-6357(00)00081-4)
- Holveck, M. J., & Riebel, K. (2014). Female zebra finches learn to prefer more than one song and from more than one tutor. *Animal Behaviour*, 88, 125–135. <https://doi.org/10.1016/j.anbehav.2013.11.023>
- Honarmand, M., Riebel, K., & Naguib, M. (2015). Nutrition and peer group composition in early adolescence: impacts on male song and female preference in zebra finches. *Animal Behaviour*, 107, 147–158. <https://doi.org/10.1016/j.anbehav.2015.06.017>
- Houx, B. B., & ten Cate, C. (1998). Do contingencies with tutor behaviour influence song learning in zebra finches? *Behaviour*, 135(5), 599–614.
- Houx, B. B., & ten Cate, C. (1999a). Do stimulus-stimulus contingencies affect song learning in zebra finches (*Taeniopygia guttata*)? *Journal of Comparative Psychology*, 113(3), 235–242. <https://doi.org/10.1037/0735-7036.113.3.235>
- Houx, B. B., & ten Cate, C. (1999b). Song learning from playback in zebra finches: is there an effect of operant contingency? *Animal Behaviour*, 57(4), 837–845. <https://doi.org/10.1006/anbe.1998.1046>
- Houx, B., Feuth, E., & ten Cate, C. (2000). Variations in zebra finch song copying: an examination of the relationship with tutor song quality and pupil behaviour. *Behaviour*, 137, 1377–1389. <https://doi.org/10.1163/156853900501980>
- Hultsch, H., Schleuss, F., & Todt, D. (1999). Auditory-visual stimulus pairing enhances perceptual learning in a songbird. *Animal Behaviour*, 58, 143–149. <https://doi.org/10.1006/anbe.1999.1120>
- Hyland Bruno, J., & Tchernichovski, O. (2019). Regularities in zebra finch song beyond the repeated motif. *Behavioural Processes*, 163, 53–59. <https://doi.org/10.1016/j.beproc.2019.07.007>

- org/10.1016/j.beproc.2017.11.001
- Immelmann, K. (1969). Song development in the zebra finch and other estrildid finches. In R. A. Hinde (Ed.), *Bird vocalizations*. Cambridge, England: Cambridge University Press,.
- Jesse, F., & Riebel, K. (2012). Social facilitation of male song by male and female conspecifics in the zebra finch, *Taeniopygia guttata*. *Behavioural Processes*, 91(3), 262–266. <https://doi.org/10.1016/j.beproc.2012.09.006>
- Jones, A. E., & Slater, P. J. B. (1993). Do young male zebra finches prefer to learn songs that are familiar to females with which they are housed. *Animal Behaviour*, 46, 616–617. <https://doi.org/10.1006/anbe.1993.1233>
- Jones, A. E., ten Cate, C., & Slater, P. J. B. (1996). Early experience and plasticity of song in adult male zebra finches (*Taeniopygia guttata*). *Journal of Comparative Psychology*, 110(4), 354–369. <https://doi.org/10.1037/0735-7036.110.4.354>
- Kojima, S., & Doupe, A. J. (2011). Social performance reveals unexpected vocal competency in young songbirds. *Proceedings of the National Academy of Sciences of the United States of America*, 108(4), 1687–1692. <https://doi.org/10.1073/pnas.1010502108>
- Krause, E. T., Bischof, H. J., Engel, K., Golüke, S., Maraci, Ö., Mayer, U., ... Caspers, B. A. (2018). Olfaction in the zebra finch (*Taeniopygia guttata*): what is known and further perspectives. *Advances in the Study of Behavior*, 50, 37–85. <https://doi.org/10.1016/bs.asb.2017.11.001>
- Krebs, J. R., & Kroodsma, D. E. (1980). Repertoires and geographical variation in bird song. *Advances in the Study of Behavior*, 11, 143–177.
- Kroodsma, D. E., & Pickert, R. (1984). Sensitive phases for song learning: Effects of social interaction and individual variation. *Animal Behaviour*, 32(2), 389–394. [https://doi.org/10.1016/S0003-3472\(84\)80274-2](https://doi.org/10.1016/S0003-3472(84)80274-2)
- Kroodsma, D. E., & Verner, J. (1978). Complex singing behaviors among *Cistothorus* wrens. *The Auk*, 95(4), 703–716.
- Leavell, B. C., Rubin, J. J., McClure, C. J. W., Miner, K. A., Branham, M. A., & Barber, J. R. (2018). Fireflies thwart bat attack with multisensory warnings. *Science Advances*, 4(8). <https://doi.org/10.1126/sciadv.aat6601>
- Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences of the United States of America*, 109(5), 1431–1436. <https://doi.org/10.1073/pnas.1114783109>
- Ljubičić, I., Hyland Bruno, J., & Tchernichovski, O. (2016). Social influences on song learning. *Current Opinion in Behavioral Sciences*, 7, 101–107. <https://doi.org/10.1016/j.cobeha.2015.12.006>
- Mann, N. I., & Slater, P. J. B. (1995). Song tutor choice by zebra finches in aviaries. *Animal Behaviour*, 49(3), 811–820. [https://doi.org/10.1016/0003-3472\(95\)80212-6](https://doi.org/10.1016/0003-3472(95)80212-6)
- Mann, N. I., Slater, P. J. B., Eales, L. A., & Richards, C. (1991). The influence of visual stimuli on song tutor choice in the zebra finch, *Taeniopygia guttata*. *Animal*

- Behaviour, 42(2), 285–293. [https://doi.org/10.1016/S0003-3472\(05\)80560-3](https://doi.org/10.1016/S0003-3472(05)80560-3)
- Mello, C. V. (2014). The zebra finch, *Taeniopygia guttata*: An avian model for investigating the neurobiological basis of vocal learning. Cold Spring Harbor Protocols, 2014(12), 1237–1242. <https://doi.org/10.1101/pdb.emo084574>
- Mennill, D. J., Newman, A. E. M., Thomas, I. P., Woodworth, B. K., Norris, D. R., Doucet, M., ... Thomas, I. P. (2018). Wild birds learn songs from experimental vocal tutors. *Current Biology*, 28(20), 3273–3278.e4. <https://doi.org/10.1016/j.cub.2018.08.011>
- Morris, D. (1954). The reproductive behaviour of the zebra finch (*Poephila Gutta-ta*), with special reference to pseudofemale behaviour and displacement activities. *Behaviour*, 6(1), 271–322. <https://doi.org/10.1163/156853954X00130>
- Morrison, R. G., & Nottebohm, F. (1993). Role of a telencephalic nucleus in the delayed song learning of socially isolated zebra finches. *Journal of Neurobiology*, 24(8), 1045–1064.
- Nelson, D. (1997). Social interaction and sensitive phases for song learning: A critical review. In C. T. Snowdon & M. Hausberger (Eds.), *Social influences on vocal development* (pp. 7–22). Cambridge, Cambridge University Press.
- Nelson, D. (1998). External validity and experimental design: the sensitive phase for song learning. *Animal Behaviour*, 56(2), 487–491. <https://doi.org/10.1006/anbe.1998.0805>
- Ohms, V. R., Snelderwaard, P. C., ten Cate, C., & Beckers, G. J. L. (2010). Vocal tract articulation in zebra finches. *PLoS ONE*, 5(7). <https://doi.org/10.1371/journal.pone.0011923>
- Oliveira, R. F., Rosenthal, G. G., Schlupp, I., McGregor, P. K., Cuthill, I. C., Endler, J. A., ... Waas, J. R. (2000). Considerations on the use of video playbacks as visual stimuli: the Lisbon workshop consensus. *Acta Ethologica*, 3(1), 61–65. <https://doi.org/10.1007/s102110000019>
- Ota, N., Gahr, M., & Soma, M. (2015). Tap dancing birds: The multimodal mutual courtship display of males and females in a socially monogamous songbird. *Scientific Reports*, 5(16614). <https://doi.org/doi:10.1038/srep16614>
- Partan, S., & Marler, P. (1999). Communication goes multimodal. *Science*, 283, 1272–1274. <https://doi.org/10.1126/science.283.5406.1272>
- Phan, M. L., Pytte, C. L., & Vicario, D. S. (2006). Early auditory experience generates long-lasting memories that may subserve vocal learning in songbirds. *Proceedings of the National Academy of Sciences of the United States of America*, 103(4), 1088–1093. <https://doi.org/10.1073/pnas.0510136103>
- Price, P. H. (1979). Developmental determinants of structure in zebra finch song. *Journal of Comparative and Physiological Psychology*, 93(2), 260–277. <https://doi.org/10.1037/h0077553>
- Pytte, C. L., & Suthers, R. A. (2000). Sensitive period for sensorimotor integration during vocal motor learning. *Journal of Neurobiology*, 42(2), 172–189. [https://doi.org/10.1002/\(SICI\)1097-](https://doi.org/10.1002/(SICI)1097-)

- 4695(20000205)42:2<172::AID-NEU2>3.0.CO;2-I
- Rek, P., & Magrath, R. D. (2016). Multimodal duetting in magpie-larks: how do vocal and visual components contribute to a cooperative signal's function? *Animal Behaviour*, 117, 35–42. <https://doi.org/10.1016/j.anbehav.2016.04.024>
- Riebel, K., Lachlan, R. F., & Slater, P. J. B. (2015). Learning and cultural transmission in chaffinch song. *Advances in the Study of Behavior*, 47, 181–227. <https://doi.org/10.1016/bs.asb.2015.01.001>
- Riebel, K., Smallegange, I. M., Terpstra, N. J., & Bolhuis, J. J. (2002). Sexual equality in zebra finch song preference: evidence for a dissociation between song recognition and production learning. *Proceedings of the Royal Society of London. Series B: Biological Sciences.*, 269(1492), 729–733. <https://doi.org/10.1098/rspb.2001.1930>
- Riebel, K. (2000). Early exposure leads to repeatable preferences for male song in female zebra finches. *Proceedings of the Royal Society of London. Series B: Biological Sciences.*, 267(1461), 2553–2558. <https://doi.org/10.1098/rspb.2000.1320>
- Riebel, K., Odom, K. J., Langmore, N. E., & Hall, M. L. (2019). New insights from female bird song: towards an integrated approach to studying male and female communication roles. *Biology Letters*, 15(4), 1–7. <https://doi.org/10.1098/rsbl.2019.0059>
- Roper, A., & Zann, R. (2006). The onset of song learning and song tutor selection in fledgling zebra finches. *Ethology*, 112(5), 458–470. <https://doi.org/10.1111/j.1439-0310.2005.01169.x>
- Rowe, C. (1999). Receiver psychology and evolution of multicomponent signals. *Animal Behaviour*, 58, 921–931. <https://doi.org/10.1006/anbe.1999.1242>
- Simon, R., Varkevisser, J., Mendoza, E., Hochradel, K., Scharff, C., Riebel, K., & Halfwerk, W. (2019). Development and application of a robotic zebra finch (RoboFinch) to study multimodal cues in vocal communication. *PeerJ Preprints* 7:E28004v3. <https://doi.org/10.7287/peerj.preprints.28004v1>
- Slater, P. J. B., Eales, L. A., & Clayton, N. S. (1988). Song learning in zebra finches (*Taeniopygia guttata*): progress and prospects. *Advances in the Study of Behaviour*, 18, 1–34. [https://doi.org/10.1016/S0065-3454\(08\)60308-3](https://doi.org/10.1016/S0065-3454(08)60308-3)
- Slater, P. J. B., & Richards, C. (1990). Renesting and song learning in the zebra finch, *Taeniopygia guttata*. *Animal Behaviour*, 40(6), 1191–1192. [https://doi.org/10.1016/S0003-3472\(05\)80190-3](https://doi.org/10.1016/S0003-3472(05)80190-3)
- Soha, J. A., & Marler, P. (2000). A species-specific acoustic cue for selective song learning in the white-crowned sparrow. *Animal Behaviour*, 60(3), 297–306. <https://doi.org/10.1006/anbe.2000.1499>
- Soha, J. A., & Peters, S. (2015). Vocal Learning in Songbirds and Humans: A Retrospective in Honor of Peter Marler. *Ethology*, 121(10), 933–945. <https://doi.org/10.1111/eth.12415>
- Soma, M. F. (2011). Social factors in song learning: a review of Estrildid finch research. *Ornithological Science*, 10(2), 89–100. <https://doi.org/10.2326/osj.10.89>

- Sossinka, R., & Böhner, J. (1980). Song types in the zebra finch. *Zeitschrift Für Tierpsychologie*, 53, 123–132. <https://doi.org/10.1111/j.1439-0310.1980.tb01044.x>
- Sturdy, C. B., & Nicoladis, E. (2017). How much of language acquisition does operant conditioning explain? *Frontiers in Psychology*, 8(OCT), 1–5. <https://doi.org/10.3389/fpsyg.2017.01918>
- Sturdy, C. B., Phillmore, L. S., Sartor, J. J., & Weisman, R. G. (2001). Reduced social contact causes auditory perceptual deficits in zebra finches, *Taeniopygia guttata*. *Animal Behaviour*, 62(6), 1207–1218. <https://doi.org/10.1006/anbe.2001.1864>
- Tanaka, M., Sun, F., Li, Y., & Mooney, R. (2018). A mesocortical dopamine circuit enables the cultural transmission of vocal behaviour. *Nature*, 563(7729), 117–120. <https://doi.org/10.1038/s41586-018-0636-7>
- Tchernichovski, O., Eisenberg-Edidin, S., & Jarvis, E. (2021). Balanced imitation sustains song culture in zebra finches. *Nature Communications*, 1–21. <https://doi.org/10.1038/s41467-021-22852-3>
- Tchernichovski, O., Mitra, P. P., Lints, T., & Nottebohm, F. (2001). Dynamics of the vocal imitation process: how a zebra finch learns its song. *Science*, 291(5513), 2564–2569. <https://doi.org/10.1126/science.1058522>
- Tchernichovski, Ofer, Nottebohm, F., Ho, C. E., Pesaran, B., & Mitra, P. P. (2000). A procedure for an automated measurement of song similarity. *Animal Behaviour*, 59(6), 1167–1176. <https://doi.org/10.1006/anbe.1999.1416>
- Tedore, C., & Johnsen, S. (2017). Using RGB displays to portray color realistic imagery to animal eyes. *Current Zoology*, 63, 27–34. <https://doi.org/10.1093/cz/zow076>
- Teinonen, T., Aslin, R. N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition*, 108(3), 850–855. <https://doi.org/10.1016/j.cognition.2008.05.009>
- ten Cate, C. (1991). Behaviour-contingent exposure to taped song and zebra finch song learning. *Animal Behaviour*, 42(5), 857–859. [https://doi.org/10.1016/S0003-3472\(05\)80131-9](https://doi.org/10.1016/S0003-3472(05)80131-9)
- ten Cate, C. (1994). Perceptual mechanisms in imprinting and song learning. In J. A. H. & J. J. Bolhuis (Ed.), *Causal mechanisms of behavioural development*. (pp. 116–146). Cambridge University Press.
- Thielcke, G. (1984). Gesangslernen beim Gartenbaumläufer (*Certhia brachydactyla*). *Die Vogelwarte*, 32, 282–297.
- Thorpe, W. H. (1954). The process of song learning in the chaffinch as studied by means of the sound spectrograph. *Nature*, 173, 465–469.
- Todt, D., Hultsch, H., & Heike, D. (1979). Conditions affecting song acquisition in nightingales (*Luscinia megarhynchos L.*). *Zeitschrift Für Tierpsychologie*, 51(1), 23–35. <https://doi.org/10.1111/j.1439-0310.1979.tb00668.x>
- Ullrich, R., Norton, P., & Scharff, C. (2016). Waltzing *Taeniopygia*: integration of courtship song and dance in the domesticated Australian zebra finch. *Animal Behaviour*, 112, 285–300. <https://doi.org/10.1016/j.anbehav.2015.11.012>

- van Kampen, H. S., & Bolhuis, J. J. (1991). Auditory learning and filial imprinting in the chick. *Behaviour*, 117, 303–319. <https://doi.org/10.1163/156853991X00607>
- van Kampen, H. S., & Bolhuis, J. J. (1993). Interaction between auditory and visual learning during filial imprinting. *Animal Behaviour*, 45, 623–625. <https://doi.org/10.1006/anbe.1993.1074>
- Waser, M. S., & Marler, P. (1977). Song learning in canaries. *Journal of Comparative and Physiological Psychology*, 91, 1–7.
- Whiten, A. (2021). The burgeoning reach of animal culture. *Science*, 372(6537). <https://doi.org/10.1126/SCIENCE.ABE6514>
- Williams, H. (1990). Models for song learning in the zebra finch: fathers or others? *Animal Behaviour*, 39(4), 745–757. [https://doi.org/10.1016/S0003-3472\(05\)80386-0](https://doi.org/10.1016/S0003-3472(05)80386-0)
- Williams, H. (2001). Choreography of song, dance and beak movements in the zebra finch (*Taeniopygia guttata*). *The Journal of Experimental Biology*, 204, 3497–3506.
- Williams, H. (2004). Birdsong and singing behavior. *Annals of the New York Academy of Sciences*, 1016, 1–30. <https://doi.org/10.1196/annals.1298.029>
- Yanagihara, S., & Yazaki-Sugiyama, Y. (2019). Social interaction with a tutor modulates responsiveness of specific auditory neurons in juvenile zebra finches. *Behavioural Processes*, 163, 32–36. <https://doi.org/10.1016/j.beproc.2018.04.003>
- Young, G. S., Merin, N., Rogers, S. J., & Ozonoff, S. (2009). Gaze behavior and affect at 6 months: Predicting clinical outcomes and language development in typically developing infants and infants at risk for autism. *Developmental Science*, 12(5), 798–814. <https://doi.org/10.1111/j.1467-7687.2009.00833.x>

