



OPEN

Concordance of movements and songs enhances receiver responses to multimodal display in the starling

Stanisław Rusiecki & Paweł Ręka

Many animals produce signals that consist of vocalizations and movements to attract mates or deter rivals. We usually consider them as components of a single multimodal signal because they are temporally coordinated. Sometimes, however, this relationship takes on a more complex spatiotemporal character, resembling choreographed music. Timing is important for audio-visual integration, but choreographic concordance requires even more skill and competence from the signaller. Concordance should therefore have a strong impact on receivers; however, little is known about its role in audio-visual perception during natural interactions. We studied the effects of movement and song type concordance in audio-visual displays of the starling, *Sturnus vulgaris*. Starlings produce two types of movements that naturally appear in specific phrases of songs with a similar temporal structure and amplitude. In an experiment with a taxidermic robotic model, males responded more to concordant audio-visual displays, which are also naturally preferred, than to discordant displays. In contrast, the effect of concordance was independent of the specific combination of movement and song types in a display. Our results indicate that the concordance of movements and songs was critical to the efficacy of the display and suggest that the information that birds gained from concordance could not be obtained by adding information from movements and songs.

Animal acoustic signals are often accompanied by body movements. In its simplest form, this relationship can be mechanistic, such that movement itself produces sound. For example, birds, bats, and insects produce sonations using movements of their limbs or wings¹. Body movements can also accompany sound signals more accidentally. Such an interaction takes place, for example, during migration, when individuals call to each other to maintain contact^{2,3}. Often, however, body movements and sounds are components of complex multimodal displays. Then, both modalities have signalling functions and are functionally related. For example, male lance-tailed manakins (*Chiroxiphia lanceolata*) court females by combining vocalizations with dance movements⁴, while male little brown frogs (*Micrixalus saxicola*) call and perform foot flagging displays in male–male agonistic interactions⁵. In some cases, such as in duetting species, the functional relationship between movements and songs is so strong that independent components are very rare, even if they can be produced separately⁶.

In audio-visual signals, body movements influence responses to songs, and this effect may take many forms. When visual and acoustic components of multimodal signals produce qualitatively similar effects separately, their joint effect can be characterized by an additive or even superadditive intensity of the reaction. European robins (*Erithacus rubecula*), for example, respond more to red-breasted singing models than to brown-breasted singing models or red-breasted quiet models⁷. Similarly, Eastern gray squirrels (*Sciurus carolinensis*) react more strongly to alarm signals consisting of the bark and tail flag than to independent components⁸. Sometimes isolated movements are rare and do not elicit responses themselves but affect responses when interacting with singing. Male blue–black grassquits (*Volatinia jacaranda*), for example, leap out of the grass to increase the detectability of the call⁹. Movements may thus work as amplifiers or attention grabbers to increase detection or discrimination of the acoustic component¹⁰. Furthermore, movements can qualitatively modulate the function of the acoustic signals. In duetting magpie-larks (*Grallina cyanoleuca*), for example, the syndrome of correlated territorial behaviours in response to the intruder's songs differs from the syndrome in response to the songs accompanied by movements⁶. Finally, singing in interaction with movement can elicit a different response than

Department of Behavioural Ecology, Faculty of Biology, Institute of Environmental Biology, Adam Mickiewicz University, Poznań, Poland. ✉email: rek@amu.edu.pl

singing itself or elicit a response only in interaction with the movement. For example, in Túngara frogs, only audio-visual signals of males attract females—isolated components are ignored¹¹.

In addition to the sounds and movements themselves, the responses to the audio-visual signal may depend on the specific match of components. In Túngara frogs, the common effect of the movements of the resonator sac and respective calls weakens with a decrease in temporal coordination¹², which suggests that the timing of both components is crucial for their integration. A similar effect of audio-visual uncoordination was observed in magpie-lark duetting displays¹³, but in these birds, movement is not necessary to elicit a response⁶. This therefore suggests that a lack of temporal coordination can sometimes have even worse consequences for the efficacy of a multimodal signal than the lack of one component. Furthermore, in some species, responses to the audio-visual signal may depend on matching certain types of movement and sound. For example, in the superb lyrebird (*Menura novaehollandiae*), song types match specific dance movements to create potent mating displays¹⁴. Chimpanzees, in turn, can combine various vocalizations with different gestures that together prompt different reactions in receivers¹⁵. Finally, in some species, such as Montezuma oropendola (*Psarocolius montezuma*), individuals adjust acoustic and visual features of their display in a way that the loudest part of their song is matched with the high range acrobatic movements and more quiet songs with less distinct body movements¹⁶. In this case, the timing is coarse, and the match concerns the amplitude of calls and range of motion, not at the fine scale of following notes but at the entire display. This concordance is therefore not mechanistic, but it is also not accidental. Nevertheless, apart from early research into the role of auditory target loudness and visual attractor size in human multimodal perception^{17,18}, little is known about the role of such choreographic concordance in communication.

We studied the relationship between sounds and movements in multimodal displays of the starling (*Sturnus vulgaris*). Starling is a medium-sized hole-nesting passerine characterized by high sociality and a complex reproductive system¹⁹. In groups, males engage in numerous interactions, establishing specific dominance hierarchies that operate in roosting and breeding areas, and males are often polygynous, defending additional breeding sites^{20,21}. Usually, males compete with their displays, countersinging and flapping their wings intensely^{22–24}. Males perform two distinct types of movements while singing their elaborated songs. Both types of movements are tightly linked with the song, only exceptionally being produced without a sound^{19,22}. At the same time, each type of movement usually accompanies a specific, structurally concordant song type; expressive wing-waving is combined mainly with conspicuous high-frequency phrases, whereas indistinct wing-flicks are typically associated with quiet rattle songs²². Due to the complexity of starling songs, such matching requires skill and competence on the part of the sender. Therefore, a correct match should give the sender an advantage in interactions with rivals or at least improve its perceived quality. If songs and wing movements arranged concordantly enhance the efficacy of the display, then the concordance should potentially be of great importance in the complex social interactions of starlings and translate into the fitness of the signaller.

We used the robotic bird model and acoustic playback to experimentally test whether the natural concordance of movement and song types affects the efficacy of audio-visual signals. We compared the responses of adult starlings to audio-visual displays consisting of concordant movements and songs (Waving + High-frequency and Flick + Rattle) and discordant movements and songs (Waving + Rattle and Flick + High-frequency) to test two predictions. First, we predicted that starlings should react more to concordant than to discordant signals. Second, we predicted that as long as the display is concordant or discordant, the responses should be independent of the specific combination of movement and song type.

Results

Starlings responded to the concordance of movement and song types. Overall, there was an effect of playback treatment on all measures of response (Fig. 1, Table 1). This effect, however, resulted mainly from a stronger reaction of birds to concordant than to discordant stimuli (Fig. 1, Table 1) and not to differences between both concordant treatments (Waving + High-frequency vs. Flick + Rattle) or between both discordant treatments (Waving + Rattle vs. Flick + High-frequency) (Fig. 1; treatment nested in concordance—Table 1). Furthermore, pairwise comparisons of responses to treatment were significant only when we compared concordant versus discordant displays (Table 2), and we did not find any significant differences between the responses to two concordant treatments or two discordant treatments (Table 2).

Discussion

Our experiment showed that displays with concordant movements and songs prompted stronger responses than discordant displays. Starlings stayed longer near the robotic model, approached the model closer, and were more likely to follow the model when the movements matched songs as in natural displays. These results indicate that starlings were able to distinguish natural pairs of songs and movements from unnatural pairs. In contrast, the concordance effect was independent of the specific combination of movements and songs in the audio-visual display. These results suggest, therefore, that the display components themselves were less important for receivers than how the two modalities were juxtaposed with each other. Overall, we suggest that the positive effect of concordance in the starling audio-visual displays is a consequence of the fact that both components together represent a multimodal percept²⁵. In other words, the receivers obtain some information only if they have access to a specific combination of both components.

Birds responded more strongly to the typical combinations of movements and songs than to the atypical combinations. Starlings most often produce both modalities in specific combinations, but this is not obligatory, as each type of movement can occur in any phrase of the song and independently^{19,23}. Therefore, there are probably no anatomical or physiological constraints on the sender's side to produce specific combinations. Nevertheless, even if such a constraint existed, it would not clearly explain the reactions of receivers. In fact,

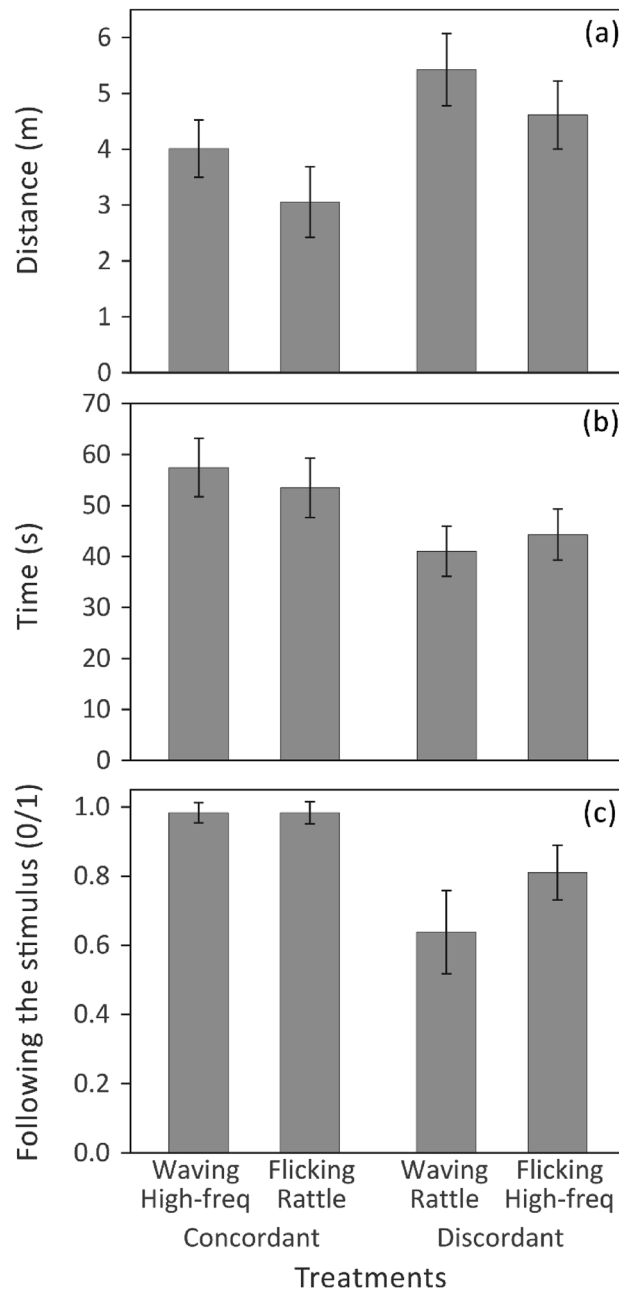


Figure 1. Response of wild starlings to four combinations of movement and song types displayed by the robotic model. Shown are responses to high-frequency and rattle song phrases accompanied by wing waving or flicking movements for 3 different response variables: (a) the minimum distance to the taxidermic model, (b) time spent by male starlings near the taxidermic model, and (c) the following of the audio-visual stimulus. Boxes show means \pm se.

Effect (df)	Distance		Time		Following the stimulus	
	F	P	F	P	F	P
Intercept (3, 70) *	2.89	0.042	3.42	0.022	3.59	0.018
Treatment (3, 70)	2.89	0.042	3.42	0.022	3.59	0.018
Intercept (3, 70) **	2.89	0.042	3.42	0.022	3.59	0.018
Concordance (1, 70)	6.08	0.016	8.50	0.005	7.42	0.008
Treatment (Concordance) (2, 70)	1.59	0.211	0.50	0.607	1.49	0.232

Table 1. The effect of treatment and audio-visual concordance on three response variables. MIXED models with either treatment effect * or concordance and treatment nested in concordance effects **.

Treatments compared		Response variables – P		
		Distance	Time	Following the stimulus
<i>Waving + High-frequency</i>	<i>Flicking + Rattle</i>	0.182	0.411	1
<i>Waving + High-frequency</i>	Waving + Rattle	0.064	0.005	0.003
<i>Waving + High-frequency</i>	Flicking + High-frequency	0.572	0.027	0.226
<i>Flicking + Rattle</i>	Waving + Rattle	0.004	0.062	0.003
<i>Flicking + Rattle</i>	Flicking + High-frequency	0.074	0.187	0.231
Waving + Rattle	Flicking + High-frequency	0.214	0.570	0.094

Table 2. Pairwise comparisons between treatment means (Fisher's LSD test). Concordant treatments are italicised. Significant values are in [bold].

the observed effect on the receiver's side may be due to the negative impact of discordant stimuli on perception. Many studies have shown that the discrepancy between sensory modalities leads to crossmodal conflicts and depression of responses²⁶. Considering our results, we think that both interpretations may be correct—in a concordant display, wing movements might enhance song efficacy, while in a discordant display, movements might hamper acoustic perception.

Our results suggest that the effect of concordance was independent of the specific combination of movement and song types in a display. Although both concordant treatments (Waving + High-frequency and Flicking + Rattle) and both discordant treatments (Waving + Rattle and Flicking + High-frequency) differed simultaneously in movement and song type, they prompted similar responses. This suggests that as long as the movement and song were concordant or discordant within a single display, it no longer mattered which particular pair of components produced the effect. In contrast, concordant and discordant stimuli always had one component in common, yet responses to them differed significantly. The fact that the difference of one component produced a stronger effect than the difference of two components also indicates that the concordance is not based on the selection of specific components for the display but on the selection of both components of the display at once based on their natural pairing. It is important to emphasize here that our experiment did not test the effect of isolated songs or movements but the effect of specific combinations of modalities. To achieve this, we standardized the intensity of both song types. However, under natural circumstances, high-frequency phrases are produced with a higher intensity than rattle phrases²⁷. Hence, we cannot exclude that in real interactions, displays with high-frequency phrases might elicit stronger responses, both in concordant and discordant displays. Nevertheless, we think that this effect would still be independent of the concordance itself.

In pairwise comparisons, one caveat was that not all concordant stimuli differed significantly from all discordant stimuli. We showed that Flicking + Rattle concordant stimuli prompted significantly stronger responses than Waving + Rattle discordant stimuli, although not significantly stronger than Flicking + High-frequency discordant stimuli. We think, however, that this discrepancy in results does not contradict our conclusions. Although some of the post-hoc (a-posteriori) tests do not confirm our main hypothesis, none of them contradicts it. At the same time, our main hypothesis does not refer to any of these pairwise comparisons, but to the a-priori contrast between concordant and discordant stimuli—and for this general comparison we obtained results that were significant for each of the tested variables (Results). The discrepancies in the results of post-hoc tests may also suggest that our study was underpowered. We did not conduct an a-priori power analysis because we had no basis for predicting the effect size of concordance. Therefore, we relied on a rule of thumb based on the Central Limit Theorem. In experimental field ecology research, the rule is that a sample size of at least 30 is safe²⁸, and we more than doubled that number. Our sample size should therefore be sufficient to obtain statistical significance for a scientifically significant effect, without overpowering the study and detecting trivial effects.

Our results imply that the study of how modalities are perceived and integrated is a necessary step to understanding the function of multimodal signals in general. Research on signalling in the sexual context is mostly unimodal. Where in turn, multimodal signalling is studied, the main emphasis is placed either on function^{29,30} or perceptual mechanisms^{31–35}. However, understanding the function of a multimodal signal requires taking into account perceptual mechanisms, as indicated by a growing number of studies^{36–39}. For example, by combining sound and motion in different structural and temporal configurations, we obtain not one multimodal signal but a whole set of signals. Depending on how the two modalities are combined, movement added to singing can enhance or weaken reactivity, or it can evoke completely different reactions^{11,13,40–44}. The concordance effect we describe here is a consequence of multimodal perception, which means that it is the specific convolution of the components that is responsible for the response. The results of this and similar experiments therefore indicate that disregarding mechanisms of multimodal perception can lead to erroneous or false negative conclusions about the function of multimodal signals²⁵.

The concordance of the starling movements and songs may provide the same basis for multimodal perception as the temporal and spatial congruence of components⁴². In humans, music is usually used as the basis for improvisation by a dancer who imitates sounds with movement because music and movements share a common structure that affords equivalent emotional expressions⁴⁵. This relationship is not limited to performance alone, as the structures responsible for musical and visual processing are cognitively related as well⁴⁶. Some animal studies have also shown that body movements can be choreographically linked to simultaneous sounds^{14,16,47}. At the same time, even in distantly related birds, brain systems that control vocal learning are directly adjacent

to brain systems involved in movement control⁴⁸, providing compelling support for a causal link between the capabilities for vocal imitation and dance. Starlings are open learners and great imitators of sounds of other species^{49,50}. If specific movements are choreographically matched with specific song phrases, this might be a strong argument for their perceptual relationship.

Materials and methods

Study site and species

We studied starlings inhabiting Szczytnicki Park in Wrocław, Poland. This area of over 1 km² is covered with a large number of old trees and lawns, providing a suitable habitat for starlings, which breed there at a high density⁵¹. After spring arrival, males start occupying nest holes and intensively display^{20,23,52}. Male starlings sing in the presence of both females and males^{20,23,52,53}. Although males are more likely to approach singing males than females^{24,54}, they are rarely aggressive and often sing alongside each other^{21,53,55}.

Starlings produce long and complex songs that typically include four categories of phrases (series of notes), carried out in a predictable sequence^{19,56}. Songs usually start with whistles, i.e., loud, pure-tone sounds, spaced by approximately 1 s pauses⁵⁶. Whistles are followed by variable phrases, which contain a variety of complex notes of relatively low amplitude and with only short temporal gaps in between. Many such notes cover a wide frequency range (1–8 kHz) in a short time, giving the impression of a series of clicks, buzzes and trills²⁷. Next, rattle phrases are sung, which are characterized by a rapid series of clicks (so-called click trains) running through the phrase as other sounds are being produced. A song bout is completed by the relatively loudest high-frequency phrases (6–10 kHz)²⁷.

Male songs are often accompanied by two kinds of wing movements, which differ both in the range of motion and temporal organization. During wing waving, half open wings rotate around shoulders (Supplementary Videos 1), while during wing flicking, wings are slightly lifted in rapid flushes (Supplementary Videos 2)¹⁹. Earlier observations showed that movements accompany 60% of song bouts; flicking was observed in 73% and waving in 33% of song bouts accompanied by movements. At the same time, 86% of wing flicks were performed during rattle phrases of songs, and 68% of wing waves were performed during terminal high-frequency phrases²². Despite the structural match, the temporal structure of these songs and movements is far from compatible. While the flicks seem to be matched to the pulses of the song (if in a rattle phrase), the waves are sequences of movements that do not exactly match any phrase of the song (Fig. 2). Therefore, apart from the nonrandom juxtaposition of specific types of movement and song, the whole display is more like a song-based movement improvisation than a fixed temporal pattern (Supplementary Videos 1 & 2).

Wing-waving has also been reported to occur without song, but the frequency of this behaviour is unknown^{19,22}. Observations of captive birds have suggested that wing movements occur primarily during courtship^{22,23,57}. However, under natural conditions, it is easy to observe males moving in the presence of only other males or in the absence of any other starling.

Experimental design

We tested a total of 74 birds to determine if the concordance of movements and songs affects the efficacy of their audio-visual signals. We used a scheme with independent groups in which each experimental male was assigned to one of four treatments consisting of audio-visual displays (Fig. 2). In each of the treatments, the birds were stimulated by audio-visual displays, consisting of the movement of the wings of the robotic model and the song played over a loudspeaker. In the concordant flicking treatment (n = 17), wing flicks were combined with the rattle song, whereas in the discordant flicking treatment (n = 20), flicks were accompanied by high-frequency terminal songs. Analogically, in the concordant waving treatment (n = 17), wing waves were accompanied by high-frequency terminal songs, whereas in the discordant waving treatment (n = 20), wing waves were accompanied by rattle songs.

To synthesize the acoustic stimuli, we used recordings from 20 males, different than the birds used in the experiment. From each individual, we selected four high-quality recordings, which were used in one series of four treatments. We used four recordings from each individual because the individual's songs showed much less structural variation of the phrases than the songs of different individuals. The recordings lasted 7–9 s, two of which consisted of 9–12 high-frequency terminal trills and another two consisted of the same number of rattle phrases. These ranges were due to differences between individuals. Therefore, instead of manipulating the length of phrases coming from different males, we preferred to use sequences of similar phrases that occur naturally in songs of one male without manipulation. Both types of songs were next combined with the two types of movements to create the four treatments. The recordings were made with the Sennheiser ME66 cardioid microphone coupled with the Olympus LS-10 solid-state recorder (sampling frequency 44.1 kHz, 16 bit) and edited using Avisoft-SASLab Pro (Avisoft Bioacoustics, Berlin, Germany).

Robotic birds

We used a taxidermic robotic male starling during playback experiments. The construction was based on two digital servo-motors (KST Digital Standard Brushless Servo MS805) placed under the model bird and attached to the humeri of wings with separate transition shafts. Thanks to this mechanism, our robot could imitate the natural behaviour of starlings. Servo-motors and audio playback were controlled together by a circuit board based on the Arduino platform (Arduino Micro; www.arduino.cc). This enabled us to precisely program the timing of wing movements and vocalizations. The whole set was remote-controlled using a custom-designed radio controller based on the X-bee platform (Digi International, Minnetonka, MN, USA).

We used video recordings of natural displays to reproduce movements of the wings. To mimic wing-waving, we used the model bird with partially opened wings moving forth and back (Supplementary Video 3). In the

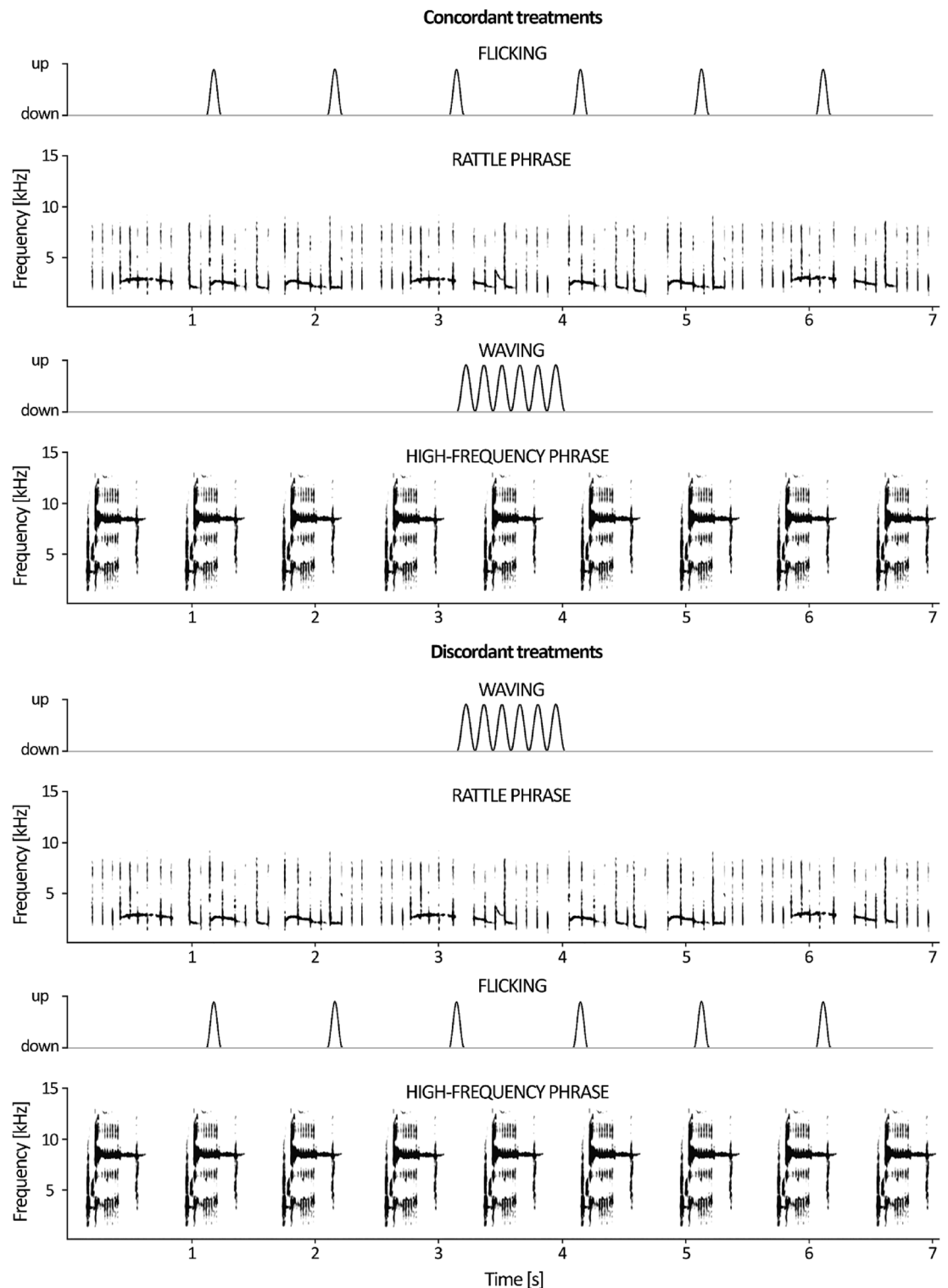


Figure 2. Combinations of visual and acoustic stimuli in treatments.

middle of each song playback, one series of 6 wing beats (movement range: 45°, duration: 1.0 s each) was given by the model. In wing-flicking treatments, the wings of the robotic bird were folded and moved up and down in 6 rapid flushes (wing beats) that lasted approximately 0.14 s each (Supplementary Video 4). During each song playback, the model performed six wing flicks in one second intervals starting one second after the playback.

Field procedures

Experiments were carried out between 23 April and 4 June 2020 in the morning hours (0600–1200 CEST). The order of treatments was randomized, and on each study day, we performed all 4 treatments with calls from one male. Before the treatment, the robotic model and the loudspeaker were mounted on the 7 m high stand to maximize its visibility. We placed the stand in the breeding sites of starlings, avoiding close proximity (less than 20 m) to nest holes. Song playbacks were broadcast from a JBL Charge 3 amplified loudspeaker (20 W, frequency range 65–20 000 Hz) at natural amplitudes of 60–65 dB SPL(A) at 10 m (measured from three individuals with UNI-T UT352 Sound Level Meter). Starlings are not territorial but social and curious, and they respond with positive phonotaxis to conspecific songs. Therefore, we lured the males near our model by playing conspicuous fragments of the starling male song and started the treatments as soon as a focal bird approached the model closer than 15 m. At the same time, we ensured that the bird had an unobstructed view of the taxidermic model. We only lured birds that were close to the given range. Therefore, the bird either immediately moved towards the model, and we started the treatment, or it ignored us, and we resigned. As a result, the luring itself had a small and standardized influence on the later reactions to the treatments. The treatment lasted one minute and included 5 randomly distributed repetitions of one song playback (7–9 s long) matched with wing movements. To avoid repeated testing of the same individuals, we accepted that the next treatment should start not earlier than 10 min after the previous treatment and at a distance of not less than 100 m. In practice, for greater certainty, we usually multiplied these values, taking into account, e.g., the direction from which the birds flew.

To compare the reactions of starlings to the treatments, we measured the time the birds were near the robot model (< 15 m), the shortest distance to the robot, and whether the birds were following the audio-visual stimulus. The time was measured from the beginning of the treatment to the moment the focal bird flew away from the model. The shortest distance to the robotic bird was initially estimated on the basis of observation and then confirmed using a laser rangefinder (Bushnell Yardage Pro). Following the stimulus refers to the movement of the bird during the treatment. After the bird had flown to less than 15 m, it either remained stationary (scored as 0) and showed no further interest in the stimulus, or it flew between the branches towards the model (scored as 1). We recorded treatments with a video camera set on a tripod 1.5 m above the ground, and we recorded audio notes in real time. Since we operated in a shaded area covered with trees, we relied primarily on direct observation and audio notes, using video recordings only for correction. We measured only males' responses to the treatments because natural starling displays are known to attract mostly same-sex conspecifics^{24,54}. We used the morphological features described in Kessel⁵⁸ and Smith et al.⁵⁹ to sex experimental birds.

Statistical analysis

We used generalized linear mixed models (MIXED) to compare bird responses to different treatments. This procedure is suitable for clustered and nonnormally distributed data. In our experiment the data was clustered because to create 74 audio playbacks we used song recordings from 20 birds. We used songs from a single male during treatments with 4 focal males, which means that the responses of such 4 focal males were not independent, and the mixed models allow for this data structure to be taken into account. We first compared responses to all four treatments. We then nested the treatment effect within the concordance effect (two concordant treatments together vs. two discordant treatments together) to test whether differences between treatments were due to concordance or differences within both these categories. Additionally, we used post hoc Fisher's LSD method to create confidence intervals and compare the means of all treatments. We separately analysed three measures of response: time, minimal distance, and following the stimulus. We fitted the models using Gaussian error distribution with logarithmic link function for time and distance variables and binomial error distribution with complementary-log-log link function for the following the stimulus variable. The identity of the bird whose song recordings were used to create the audio playbacks was used as a random factor in the models. Except for the following the stimulus variable, this random effect was significant in every analysis. This therefore confirms that the use of recordings from one individual in four different treatments was justified. For all the analyses, we used IBM SPSS Statistics v. 27.0 (Armonk, NY, USA). All *P* values were two-tailed.

Ethics statement

The experimental protocol adhered to the Animal Behaviour Society guidelines for the use of animals in research. Necessary permits were obtained from the Polish Regional Direction of Environmental Protection (WPN-II.6401.377.2017. AC, WPN.6401.345.2017.MK).

Data availability

The data reported in this paper are available in electronic supplementary material, Dataset S1.

Received: 4 December 2023; Accepted: 7 February 2024

Published online: 13 February 2024

References

- Clark, C. J. Ways that animal wings produce sound. *Integr. Comp. Biol.* **61**, 696–709. <https://doi.org/10.1093/icb/icab008> (2021).
- Farnsworth, A. Flight calls and their value for future ornithological studies and conservation research. *Auk* **122**, 733–746. <https://doi.org/10.1093/auk/122.3.733> (2005).
- Edds-Walton, P. L. Acoustic communication signals of mysticete whales. *Bioacoustics* **8**, 47–60. <https://doi.org/10.1080/09524622.1997.9753353> (1997).
- DuVal, E. H. Cooperative display and lekking behavior of the lance-tailed manakin (*Chiroxiphia lanceolata*). *Auk* **124**, 1168–1185. [https://doi.org/10.1642/0004-8038\(2007\)124\[1168:CDALBO\]2.0.CO;2](https://doi.org/10.1642/0004-8038(2007)124[1168:CDALBO]2.0.CO;2) (2007).

5. Krishna, S. & Krishna, S. Visual and acoustic communication in an endemic stream frog, *Micrixalus saxicolus* in the Western Ghats, India. *Amphibia-Reptilia* **27**, 143–147. <https://doi.org/10.1163/156853806776052056> (2006).
6. Reş, P. & Magrath, R. D. Multimodal duetting in magpie-larks: how do vocal and visual components contribute to a cooperative signal's function?. *Anim. Behav.* **117**, 35–42. <https://doi.org/10.1016/j.anbehav.2016.04.024> (2016).
7. Chantrey, D. F. & Workman, L. Song and plumage effects on aggressive display by the European Robin *Erithacus rubecula*. *Ibis* **126**, 366–371. <https://doi.org/10.1111/j.1474-919X.1984.tb00257.x> (1984).
8. Partan, S. R., Fulmer, A. G., Gounard, M. A. M. & Redmond, J. E. Multimodal alarm behavior in urban and rural gray squirrels studied by means of observation and a mechanical robot. *Curr. Zool.* **56**, 313–326. <https://doi.org/10.1093/czoolo/56.3.313> (2010).
9. Wilczynski, W., Ryan, M. J. & Brenowitz, E. A. The display of the Blue-black grassquit: The acoustic advantage of getting high. *Ethology* **80**, 218–222. <https://doi.org/10.1111/j.1439-0310.1989.tb00741.x> (1989).
10. Hebets, E. A. & Papaj, D. R. Complex signal function: Developing a framework of testable hypotheses. *Behav. Ecol. Sociobiol.* **57**, 197–214. <https://doi.org/10.1007/s00265-004-0865-7> (2005).
11. Taylor, R. C. & Ryan, M. J. Interactions of multisensory components perceptually rescue Túngara frog mating signals. *Science* **341**, 273–274. <https://doi.org/10.1126/science.1237113> (2013).
12. Taylor, R. C., Klein, B. A., Stein, J. & Ryan, M. J. Multimodal signal variation in space and time: how important is matching a signal with its signaler?. *J. Exp. Biol.* **214**, 815–820. <https://doi.org/10.1242/jeb.043638> (2011).
13. Reş, P. Multimodal coordination enhances the responses to an avian duet. *Behav. Ecol.* **29**, 411–417. <https://doi.org/10.1093/beheco/axx174> (2018).
14. Dalziell, A. H. *et al.* Dance choreography is coordinated with song repertoire in a complex avian display. *Curr. Biol.* **23**, 1132–1135. <https://doi.org/10.1016/j.cub.2013.05.018> (2013).
15. Wilke, C. *et al.* Production of and responses to unimodal and multimodal signals in wild chimpanzees *Pan troglodytes schweinfurthii*. *Anim. Behav.* **123**, 305–316. <https://doi.org/10.1016/j.anbehav.2016.10.024> (2017).
16. Miles, M. C. & Fuxjager, M. J. Animal choreography of song and dance: a case study in the Montezuma oropendola *Psarocolius montezuma*. *Anim. Behav.* **140**, 99–107 (2018).
17. Vroomen, J., Bertelson, P. & De Gelder, B. The ventriloquist effect does not depend on the direction of automatic visual attention. *Percept. Psychophys.* **63**, 651–659. <https://doi.org/10.3758/BF03194427> (2001).
18. Radeau, M. Signal intensity, task context, and auditory-visual interactions. *Perception* **14**, 571–577. <https://doi.org/10.1068/p140571> (1985).
19. Feare, C. *The Starling* (Oxford University Press, 1984).
20. Pinxten, R., Verheyen, R. F. & Eens, M. Polygyny in the European starling. *Behaviour* **111**, 234–256. <https://doi.org/10.1163/156853989X00682> (1989).
21. Ellis, C. R. Jr. Agonistic behavior in the male starling. *Wilson Bull.* **78**, 208–224 (1966).
22. Böhner, J. & Veit, F. Song structure and patterns of wing movement in the European starling (*Sturnus vulgaris*). *J. Ornithol.* **134**, 309–315. <https://doi.org/10.1007/BF01640426> (1993).
23. Eens, M., Pinxten, R. & Verheyen, R. F. On the function of singing and wing-waving in the European starling *Sturnus vulgaris*. *Bird Study* **37**, 48–52. <https://doi.org/10.1080/00063659009477038> (1990).
24. Mountjoy, D. J. & Lemon, R. E. Song as an attractant for male and female European starlings, and the influence of song complexity on their response. *Behav. Ecol. Sociobiol.* **28**, 97–100. <https://doi.org/10.1007/BF00180986> (1991).
25. Halfwerk, W. *et al.* Toward testing for multimodal perception of mating signals. *Front. Ecol. Evol.* **7**, 124. <https://doi.org/10.3389/fevo.2019.00124> (2019).
26. Talsma, D., Senkowski, D., Soto-Faraco, S. & Woldorff, M. G. The multifaceted interplay between attention and multisensory integration. *Trends Cogn. Sci.* **14**, 400–410. <https://doi.org/10.1016/j.tics.2010.06.008> (2010).
27. Eens, M., Pinxten, R. & Verheyen, R. F. Temporal and sequential organization of song bouts in the starling. *Ardea* **77**, 75–86 (1989).
28. Martínez-Abraín, A. Is the 'n = 30 rule of thumb' of ecological field studies reliable? A call for greater attention to the variability in our data. *Anim. Biodivers. Conserv.* **37**, 95–100 (2014).
29. Partan, S. R. & Marler, P. Issues in the classification of multimodal communication signals. *Am. Nat.* **166**, 231–245. <https://doi.org/10.1086/431246> (2005).
30. Candolin, U. The use of multiple cues in mate choice. *Biol. Rev.* **78**, 575–595. <https://doi.org/10.1017/s1464793103006158> (2003).
31. Whitchurch, E. A. & Takahashi, T. T. Combined auditory and visual stimuli facilitate head saccades in the Barn owl (*Tyto alba*). *J. Neurophysiol.* **96**, 730–745. <https://doi.org/10.1152/jn.00072.2006> (2006).
32. Spence, C. Crossmodal correspondences: A tutorial review. *Atten. Percept. Psychophys.* **73**, 971–995. <https://doi.org/10.3758/s13414-010-0073-7> (2011).
33. Maier, J. X., Chandrasekaran, C. & Ghazanfar, A. A. Integration of bimodal looming signals through neuronal coherence in the temporal lobe. *Curr. Biol.* **18**, 963–968. <https://doi.org/10.1016/j.cub.2008.05.043> (2008).
34. Perrodin, C., Kayser, C., Logothetis, N. K. & Petkov, C. I. Natural asynchronies in audiovisual communication signals regulate neuronal multisensory interactions in voice-sensitive cortex. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 273–278. <https://doi.org/10.1073/pnas.1412817112> (2015).
35. Chen, L. & Vroomen, J. Intersensory binding across space and time: a tutorial review. *Atten. Percept. Psychophys.* **75**, 790–811. <https://doi.org/10.3758/s13414-013-0475-4> (2013).
36. Hebets, E. A. *et al.* A systems approach to animal communication. *Proc. R. Soc. Lond. B* **283**, 20152889. <https://doi.org/10.1098/rspb.2015.2889> (2016).
37. Starnberger, I., Preininger, D. & Hödl, W. From uni- to multimodality: Towards an integrative view on anuran communication. *J. Comp. Physiol. A* **200**, 777–787. <https://doi.org/10.1007/s00359-014-0923-1> (2014).
38. Halfwerk, W. & Slabbekoorn, H. Pollution going multimodal: The complex impact of the human-altered sensory environment on animal perception and performance. *Biol. Lett.* **11**, 20141051. <https://doi.org/10.1098/rsbl.2014.1051> (2015).
39. Ryan, M. J., Page, R. A., Hunter, K. L. & Taylor, R. C. 'Crazy love': Nonlinearity and irrationality in mate choice. *Anim. Behav.* **147**, 189–198. <https://doi.org/10.1016/j.anbehav.2018.04.004> (2019).
40. Reş, P. & Magrath, R. D. Display structure size affects the production of and response to multimodal duets in magpie-larks. *Anim. Behav.* **187**, 137–146. <https://doi.org/10.1016/j.anbehav.2022.03.005> (2022).
41. Reş, P. & Magrath, R. D. Reality and illusion: The assessment of angular separation of multi-modal signalers in a duetting bird. *Proc. R. Soc. Lond. B* **289**, 20220680. <https://doi.org/10.1098/rspb.2022.0680> (2022).
42. Narins, P. M., Grabul, D. S., Soma, K. K., Gaucher, P. & Hödl, W. Cross-modal integration in a dart-poison frog. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 2425–2429. <https://doi.org/10.1073/pnas.0406407102> (2005).
43. Kozak, E. C. & Uetz, G. W. Cross-modal integration of multimodal courtship signals in a wolf spider. *Anim. Cogn.* **19**, 1173–1181. <https://doi.org/10.1007/s10071-016-1025-y> (2016).
44. Halfwerk, W., Page, R. A., Taylor, R. C., Wilson, P. S. & Ryan, M. J. Crossmodal comparisons of signal components allow for relative-distance assessment. *Curr. Biol.* **24**, 1751–1755. <https://doi.org/10.1016/j.cub.2014.05.068> (2014).
45. Sievers, B., Polansky, L., Casey, M. & Wheatley, T. Music and movement share a dynamic structure that supports universal expressions of emotion. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 70–75. <https://doi.org/10.1073/pnas.1209023110> (2013).
46. Levitin, D. J. & Tirovolas, A. K. Current advances in the cognitive neuroscience of music. *Ann. N. Y. Acad. Sci.* **1156**, 211–231. <https://doi.org/10.1111/j.1749-6632.2009.04417.x> (2009).

47. Patel, A. D., Iversen, J. R., Bregman, M. R. & Schulz, I. Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Curr. Biol.* **19**, 827–830. <https://doi.org/10.1016/j.cub.2009.03.038> (2009).
48. Feenders, G. *et al.* Molecular mapping of movement-associated areas in the avian brain: a motor theory for vocal learning origin. *PLoS ONE* **3**, e1768. <https://doi.org/10.1371/journal.pone.0001768> (2008).
49. Hindmarsh, A. M. Vocal mimicry in starlings. *Behaviour* **90**, 302–324. <https://doi.org/10.1163/156853984X00182> (1984).
50. Marler, P., Böhner, J. & Chaiken, M. Repertoire turnover and the timing of song acquisition in European starlings. *Behaviour* **128**, 25–39. <https://doi.org/10.1163/156853994X00037> (1994).
51. Tomiałojć, L. Changes in breeding bird communities of two urban parks in Wrocław across 40 years (1970–2010): Before and after colonization by important predators. *Ornis Pol.* **52**, 1–25 (2011).
52. Eens, M., Pinxten, R. & Frans, R. V. Function of the song and song repertoire in the European starling (*Sturnus vulgaris*): an aviary experiment. *Behaviour* **125**, 51–66. <https://doi.org/10.1163/156853993X00182> (1993).
53. Henry, L., Hausberger, M. & Jenkins, P. F. The use of song repertoire changes with pairing status in male European starling. *Bioacoustics* **5**, 261–266. <https://doi.org/10.1080/09524622.1994.9753256> (1994).
54. Gentner, T. Q. & Hulse, S. H. Perceptual mechanisms for individual vocal recognition in European starlings *Sturnus vulgaris*. *Anim. Behav.* **56**, 579–594. <https://doi.org/10.1006/anbe.1998.0810> (1998).
55. Hausberger, M., Richard-Yris, M. A., Henry, L., Lepage, L. & Schmidt, I. Song sharing reflects the social organization in a captive group of European starlings (*Sturnus vulgaris*). *J. Comp. Psychol.* **109**, 222–241. <https://doi.org/10.1037/0735-7036.109.3.222> (1995).
56. Eens, M. Understanding the complex song of the European starling: an integrated ethological approach. *Adv. Stud. Behav.* **26**, 355–434. [https://doi.org/10.1016/S0065-3454\(08\)60384-8](https://doi.org/10.1016/S0065-3454(08)60384-8) (1997).
57. Eens, M., Pinxten, R. & Verheyen, R. Variation in singing activity during the breeding cycle of the European starling *Sturnus vulgaris*. *Belg. J. Zool.* **124**, 167–174 (1994).
58. Kessel, B. Criteria for sexing and aging European starlings (*Sturnus vulgaris*). *Bird-Banding* **22**, 16–23. <https://doi.org/10.2307/4510224> (1951).
59. Smith, E. L. *et al.* Sexing starlings *Sturnus vulgaris* using iris colour. *Ring. Migr.* **22**, 193–197. <https://doi.org/10.1080/03078698.2005.9674332> (2005).

Acknowledgements

We thank Hanna Michalska for discussion and two anonymous reviewers for their valuable feedback, which helped to improve the manuscript. This work was supported by the National Science Centre, Poland (Grant No. 2022/45/B/NZ8/00884).

Author contributions

S.R. and P.R. conceived the study and contributed to writing; S.R. conducted the study.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-54024-w>.

Correspondence and requests for materials should be addressed to P.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024