Function of a multimodal signal: a multiple-hypotheses test using an electromechanical robot

Vinicius Matheus Caldart[1,2] | Mauricio Beux dos Santos[2] | Glauco Machado[3]

[1] School of Biology, Centre of Biological Diversity, University of St Andrews, St Andrews, United Kingdom

[2] Departamento de Bioquímica e Biologia Molecular, Programa de Pós-graduação em Bioquímica Toxicológica, Universidade Federal de Santa Maria, Santa Maria, RS, Brazil

[3] LAGE do Departamento de Ecologia, Instituto de Biociências, Universidade de São Paulo, São Paulo, SP, Brazil

Correspondence

Vinicius Matheus Caldart

Email: viniciuscaldart@gmail.com

Running head: A multiple-hypotheses test of multimodal signal function

FUNDING INFORMATION

São Paulo Research Foundation (FAPESP grant 2015/14336-6)

National Council for Scientific and Technological Development (CNPq grant 306550/2014-8)

Abstract

1. Many animal species communicate using multimodal signals, which are composed of two or more components emitted and interpreted through different sensory modalities. The main types of selective pressures leading to the evolution of multimodal signals are: (1) *content-based*, when combined components convey information about the signaller, (2) *efficacy-based*, when combined components increase the efficacy of signal transmission or processing, and (3) *inter-signal interaction*, when combined components act in concert to elicit a receiver's response that differs from the response to either components alone.

2. We experimentally tested predictions of five competing hypotheses about the selective pressures leading to the evolution of an audiovisual display (aggressive calls + toe flags) emitted during agonistic male-male interactions of the Neotropical frog *Crossodactylus schmidti*.

3. To simulate these agonistic interactions in the field, we built an electromechanical robot that emitted acoustic and visual stimuli, either combined or in isolation. We assessed male receiver's responses to the stimuli types in terms of occurrence and frequency of signals emission.

4. The hypothesis that received most support was the *context hypothesis*, which states that one signal provides a context in which a receiver can interpret and respond to a second signal. The main findings supporting this hypothesis are: (1) the receiver's response to the acoustic and visual stimuli were non-redundant; (2) the multimodal stimulus magnified the receiver's response; and (3) the frequency of aggressive notes and toe flags covaried in the receiver's response to the multimodal stimulus.

5. We suggest that the addition of toe flags to aggressive calls modulates male-male interactions by providing to the receiver a new context for the interpretation and response to the acoustic signal. This new context may be motivation to fight for territory possession, level of aggressiveness, or intent of territory defence or invasion. Our study provides one of the few evidences of a context function of a multimodal signal with a modulation effect in animal contests. Moreover, our results draw attention to the role of inter-signal selection on the evolution of multimodal signals in species living in heterogeneous habitats, where efficacy-based selection is often considered as a major driver.

Key words: audiovisual display, complex communication, context hypothesis, inter-signal interaction, modulation, multimodal signal evolution.

1 | INTRODUCTION

In numerous taxa, individuals communicate using multimodal signals, which are characterized by always being composed of two or more components emitted and interpreted through two or more sensory modalities (Hebets & Papaj, 2005; Partan & Marler, 2005). Multimodal signals occur in both invertebrates (e.g., Hölldobler, 1999; Wilgers & Hebets, 2012) and vertebrates (e.g., Starnberger et al., 2014; Freeman & Hare, 2015), usually consisting of signals emitted simultaneously or in sequence. Although widespread in nature, we still have an incipient understanding of why multimodal communication has evolved and what are the functions of multimodal signals (Higham & Hebets, 2013). These are central questions in evolutionary biology because animal signals tend to be costly to produce, to maintain or to emit (Zahavi, 1975; Endler, 1992).

Three types of selective pressures driving the evolution of multimodal signals are recognized. In the *content-based* selection, combined components convey information about the signaller, including quality, condition, location, and species or sex identity. The 'redundant signal' hypothesis is a content-based explanation for the fact that higher quality zebra finch males have redder beaks and higher song rates. Both traits are expressed only by males in good condition, an attribute that females access more accurately when visual and acoustic signals are combined (Birkhead et al., 1998). In the *efficacy-based* selection, combined components increase the efficacy of signal transmission or processing through the environment. The 'efficacy trade-off' hypothesis is an efficacy-based explanation for the simultaneous vocal sac movements and call emission in the túngara frog because each signal is specialized to overcome a particular transmission impediment (Rosenthal et al., 2004). Finally, in the *inter-signal interaction* selection, combined components are temporally coupled and act in concert to elicit a receiver's response that differs from the response to either components alone (Hebets & Papaj, 2005). The 'context' hypothesis is an inter-signal interaction explanation for the fact that chemical signals alter a male's response to the visual display of an open chela in snapping shrimps. Males show more intense responses to the visual display in the presence of chemical signals from males rather than females, suggesting that the presence of one signal provides a context in which a receiver can interpret and respond to a second signal (Hughes, 1996).

In a review on the function of complex signals, Hebets & Papaj (2005) provided a framework of testable hypotheses and suggested that one could narrow down the possible explanations for the function of a multimodal display by addressing four key questions (Table 1): (1) Are the signals necessary or sufficient to elicit a receiver's response? (2) Does the presence of one signal influence the receiver's response to a second signal? (3) Do the signals covary? (4) Do any/all of the signals covary with signaller quality? Other theoretical contribution on the function of multimodal signals was made by Partan & Marler (2005). Based on the elicited receiver's response, the authors classified signals as redundant, when isolated components generate a qualitatively equivalent response, or non-redundant, when the components generate qualitatively different responses. They also proposed a classification of multimodal signals according to the expected receiver's response to combined signals (Table 1). The two studies mentioned above offer the opportunity of testing multiple competing hypotheses, which is a powerful yet underused way of strong inference in ecology (Betini et al., 2017). Here we used the approach of multiple competing hypotheses to understand the functional meaning of an audiovisual display emitted by a diurnal stream-breeding frog, *Crossodactylus schmidti* (Hylodidae), which occurs in the neotropics (Figure 1A).

A common audiovisual display of *C. schmidti* is composed of simultaneous emissions of aggressive calls and toe flags by males during agonistic encounters with conspecifics for the possession of calling sites (Caldart et al., 2014). Toe flags consist of up-and-down movements performed with the toes, which are lifted from the substrate, showing the contrasting coloration between their ventral and dorsal parts (Hartmann et al., 2005; see video in Appendix S1). Although frequent in the social interactions of *C. schmidti* (Caldart et al., 2014) and other anuran species living in noisy habitats, such as waterfalls or torrent streams (Hödl & Amézquita, 2001), to what extent toe flags represent a visual signal or interact with acoustic components to elicit a different receiver's response is still an open question. To answer this question, we built an electromechanical robot and programmed it to emit acoustic and visual stimuli, either combined or in isolation, in a field experiment (Figures 1B-C; Appendix S2). The data gathered in this experiment were used to test a set of competing hypotheses based on Hebets & Papaj (2005) and Partan & Marler (2005).

Here we selected *a priori* five hypotheses whose assumptions are compatible with what we know on the ecology and behaviour of *C. schmidti* (e.g., Caldart et al., 2011, 2014, 2016a,b, 2019; Appendix S3). These hypotheses are presented below and a summary with their respective unique combination of predictions is presented in Table 1. Among these hypotheses, there are two related to content-based selection (*quality* and *redundant signal*), one related to efficacy-based selection (*efficacy backup*), and two related to inter-signal interaction selection (*attention-altering* and *context*) — all of them extracted from Hebets & Papaj, (2005). The *quality* hypothesis postulates that different signals convey different aspects of signaller quality, such as size and condition. The *redundant signal* hypothesis postulates that different signals provide the same information about the signaller, allowing for an increased accuracy of receiver's response. The *efficacy backup* hypothesis postulates that each signal acts as a backup to the other in the presence of environmental variability, so that the probability of receiver's response across some environmental conditions is higher when both signals are present. The *attention-altering* hypothesis postulates that one signal influences the information filtering mechanism of the receiver, thereby decreasing the reaction time to another signal. Finally, the *context* hypothesis postulates that one signal provides a context in which a receiver can interpret and respond to a second signal. These hypotheses were tested under field conditions to disentangle the functional meaning of the acoustic, visual, and audiovisual signals emitted by male frogs in a noisy habitat, contrasting different types of selective pressures recognized for driving multimodal signal evolution in animals.

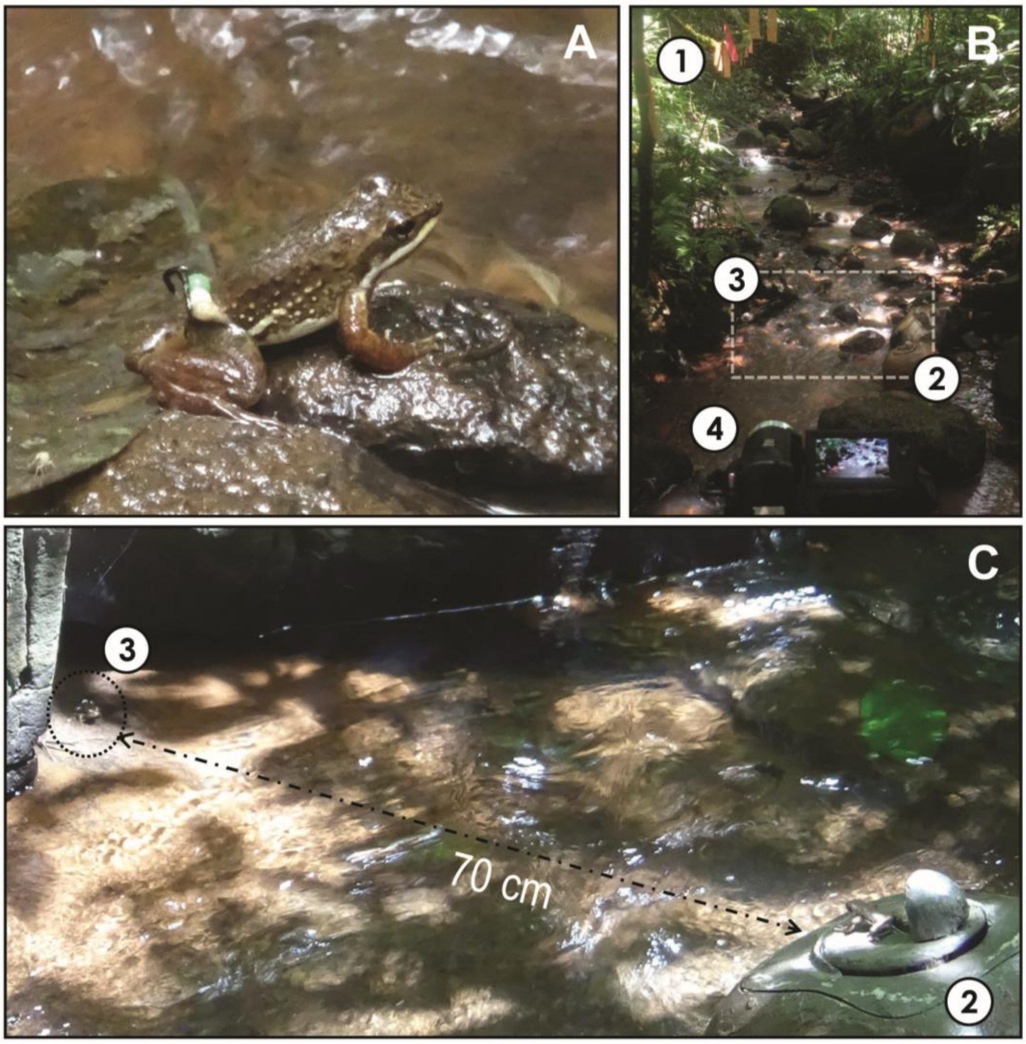

FIGURE 1 (A) Marked male of *Crossodactylus schmidti* on a rock, the typical signalling site in the torrent streams inhabited by the individuals. (B) General view of the experimental setup. In both (B) and (C) the numbers indicate: (1) coloured flags marking the territories of individually marked males; (2) the electromechanical robot (EMR); (3) a focal male individually marked with a temporary waist belt; and (4) a video-camera placed 3 m from the EMR and 3.7 m from the focal male. In all trials, the EMR was placed at a viewing distance of 70 cm from and at an angle of 30° relative to the focal male.

TABLE 1 Set of hypotheses selected *a priori* to understand the function of a multimodal signal emitted by males of the frog *Crossodactylus schmidti*. The letter "*a*" represents the *acoustic* component (aggressive call) and the letter "*v*" represents the *visual* component (toe flags). Predictions related to questions (1) and (4-6) follow Hebets & Papaj (2005), while predictions related to questions (2-3) follow Partan & Marler (2005). In Appendix S3 we present detailed explanations on the inclusion and exclusion of all hypotheses proposed by Hebets & Papaj (2005). The symbol "+" represents a positive correlation and the symbol "-" represents a negative correlation between different signals. Underlined predictions are those supported by our field experiment (see Results).

Questions	**Content-based hypotheses**		**Efficacy-based hypothesis**	**Inter-signal interaction hypothesis**	
	Quality	**Redundant signal**	**Efficacy backup**	**Attention-altering**	**Context**
(1) Is each signal (*a* and *v*) alone necessary or sufficient to elicit a receiver's response?	Yes	Yes	Yes	Yes	No
(2) Are the expected receiver's responses to signals *a* and *v* redundant or non-redundant?	Non-redundant[a]	Redundant[a]	Redundant[a]	Redundant[a]	Non-redundant[a]
(3) What is the category of the expected receiver's response to signal *a*+*v*?	Independence or dominance[b]	Enhancement[c]	Enhancement[c]	Enhancement[c]	Modulation[d]
(4) Do the signals *a* and *v* covary in the receiver's response?	No	Yes (+ or -)	Yes (+)	No	Yes (+)
(5) Do signals *a* and/or *v* in the receiver's response covary with the receiver quality?	Yes (both *a* and *v*)	Yes (both *a* and *v*)	No	No	No
(6) Is the probability of receiver's response influenced by environmental conditions and higher in response to *a*+*v*?	No	No	Yes	No	No

[a] In *non-redundant* signals, the separate components are followed by qualitatively different responses. In *redundant* signals, the separate components are followed by qualitatively equivalent responses. [b] *Independence* means that non-redundant components in combination have each an independent effect. *Dominance*, in turn, means that one component overshadows the effect of the other. [c] *Enhancement* means that the effect of redundant components in combination generates a more intense response than that of either component alone. [d] *Modulation* means that combined non-redundant signal components can minimize or magnify the effect of the other. By definition, modulation implies that the presence of signal *a* in a multimodal signal alters the receiver's response to signal *v* (or *v* to *a*), which is a prediction of the context hypothesis. Terminology based on Partan & Marler (2005).

2 | MATERIALS AND METHODS

2.1 | Study area

We conducted the study in two torrent streams at Turvo State Park (27°14′34.08″S, 53°57′13.74″W, 376 m a.s.l.), municipality of Derrubadas, state of Rio Grande do Sul, southern Brazil. The park covers an area of 17,500 ha covered by semi-deciduous forest. The local climate is characterized as subtropical sub-humid with a dry summer. Mean annual rainfall is 1,787 mm and mean annual temperature is 18.8 °C (Maluf, 2000).

2.2 | Stimulus preparation and programming

We developed an electromechanical robot (EMR) presenting the average snout-vent length of a *C. schimidti* male, i.e., 26.65 mm (Caldart et al., 2012). The structure, operation, and efficacy of the EMR will be described elsewhere, but a brief view of its functioning is shown in Appendix S2. We programmed the EMR to emit aggressive calls and toe flags, either combined (multimodal signal) and in isolation (unimodal signal). Thus, we created three experimental groups: (1) *acoustic* (only aggressive calls), (2) *visual* (only toe flags), and (3) *multimodal* (both aggressive calls and toe flags). The three experimental groups programmed in the EMR had the same temporal structure, composed of a 4 min stimulus phase preceded and followed by a 4 min control phases (pre- and post-controls), totalling 12 min. The multimodal stimulus was composed of a train of 12 s of aggressive call with a simultaneous toe flag in every 12 notes, followed by 28 s of silent interval. This sequence was repeated six times during the stimulus phase. The acoustic and the visual stimuli were composed of the aggressive call alone and the toe flags alone, respectively. The exact moments of signal exhibition during the stimulus phase were the same for the three types of stimuli (Appendix S2). During the pre- and post-control phases the EMR remained silent and motionless.

To represent a typical aggressive call and to allow it to be programmed in the EMR, we modified in Adobe Audition the parameters of a characteristic aggressive call recorded from an average-sized male used for the description of the species' calls (Caldart et al., 2011; Appendix S2). As for the visual stimulus programmed in the EMR, we selected film recordings of nine males previously analysed by Caldart et al. (2014) and counted the number of aggressive notes emitted between the emissions of two consecutive toe flags (Appendix S2).

2.3 | Field experiment

We located and captured males in calling activity along torrent streams between October and December 2016, and in January 2017, totalling 25 days of fieldwork. In these months, male activity and frequency of mature adults do not vary significantly in the population (Caldart et al., 2016a, 2019). We recorded the snout-vent length (to the nearest 0.01 mm) and the body mass (to the nearest 0.1 g) of all males and marked them with a temporary cotton waist belt around the pelvic region containing an individual code (Figure 1A). After releasing each male to his calling site, we tied a flag containing his individual code in the vegetation right above the site he was captured (Figure 1B). Because males are territorial, we could locate the same individuals during the experiment to expose each of them to all experimental groups.

One day after the marking procedure, we searched for marked males from 09:00 h to 17:00 h. We observed each marked male for 5 min and, if the male did not interact with conspecifics during this period, we positioned the EMR at a viewing distance of 70 cm, and at an angle of 30°, relative to the focal male (Figures 1B-C). Then we waited 5 min for acclimatization and, if no activity by the focal male was recorded, we exposed him to one of the three experimental groups (visual, acoustic, and multimodal). To avoid influence of the order of exposure of the types of stimulus in male responses, we randomized the sequence of stimuli presentation by the EMR to each individual.

We recorded the entire period of experiment (12 min) for each focal male using a digital camcorder (Sony Handycam HDR-CX405, Figures 1B-C) and, immediately after the footage, we measured the noise (dBC) and light intensity levels (lux) 30 cm above the calling site of the focal male, with a sound lever meter (Instrutherm DEC 500; C weighting curve: 20–12,500 Hz, dB range: 35–130 dB) and a luximeter (Instrutherm LD400), respectively. After finishing a trial, we either waited for at least 30 min (plus 5 min of acclimatization) before assigning other stimulus to the same focal male or moved the EMR to another site and repeated the procedure with other male. We recorded 39 males, from which 25 were exposed to the three types of stimulus and 14 males were exposed to two types of stimulus (acoustic + visual = 2; acoustic + multimodal = 5; visual + multimodal = 7). From the total of 39 males, we excluded one male that did not respond in any phase of any experimental group. Thus, our total sample size was 38 males, from which 33 males were exposed to the visual, 31 males were exposed to the acoustic, and 36 males were exposed to the multimodal stimulus.

2.4 | Statistical analyses

To answer the questions presented in Table 1, we used different sets of variables, which are summarized in Table 2. To encompass all the signals known to occur in male-male interactions, we used as response variables either the occurrence and/or the frequency of aggressive notes, frequency of toe flags, and frequency of other visual (non-toe flags) signals emitted by focal males, such as limb lifting, running-jumping display, and body jerking (see description of these behaviours in Caldart et al., 2014).

To answer questions (1-3) and (5-6), we performed generalized linear mixed models (GLMMs) using different families of error distribution of the response variables and the most adequate link functions (Table 2). We used the package *glmmTMB* (Brooks et al., 2017) for model construction and the package *DHARMa* (Hartig, 2020) for model evaluation. In all models we used male identity as a random factor to control for repeated exposure of the same individuals to different experimental phases and experimental groups.

To answer questions (1-3), we used the frequency of signals as continuous response variable, and the experimental phases as categorical predictors (Table 2). To test *post-hoc* differences in male responses between experimental phases of the experimental groups, we computed the estimated marginal means of the response variable for each of the experimental phases (with 95% confidence intervals) based on the fitted model, and calculated customized contrasts to compare the estimated marginal means with one another using the package *emmeans* (Lenth, 2019). We used the *mvt* alpha adjustment method for multiple contrasts from the *mvtnorm* package (Genz et al., 2020). The *mvt* is a one-step multiple comparison method that uses a multivariate *t* distribution with the same covariance structure as the model estimates to determine the p-value adjustment.

To answer question (4), we used Spearman correlation coefficients between male signals emitted during the stimulus phase of the experimental groups. We considered not only the frequency of aggressive notes and toe flags, but also the frequency of aggressive calls to test for correlations between acoustic and visual signals, because aggressive calls are highly variable in duration (Caldart et al., 2011) and it is unknown if the number of calls and number of notes emitted by the males are correlated. For the visual experimental group, we could not test correlations between signals emitted by focal males because only two of them emitted toe flags during the stimulus phase.

To answer question (5), we used as response variables the occurrence and frequency of male signals emitted during the stimulus phase of each experimental group (Table 2). We estimated body condition of the focal males as the residuals of a linear regression between

$\log_{10}$ transformed body mass and $\log_{10}$ transformed snout-vent length and used the residual values as a proxy of signaller quality (Appendix S5). Negative residual values indicate that males were in poor body condition, whereas positive values indicate that male were in good body condition (Băncilă et al., 2010). Because body size may indicate both individual age — a relationship already reported for *C. schmidti* (Caldart et al., 2019) — and individual resource holding potential (Dyson et al., 2013), we also used male size, measured as the snout-vent length, as another proxy of male quality. Both continuous predictor variables were standardized and centred to zero to make their coefficients comparable.

Finally, to answer question (6), we used the occurrence of any male response (e.g., body movements towards the EMR, calls, visual or multimodal displays) because we were interested in the probability of male response irrespective of the type of signal emitted. However, all males showed at least one type of response during the 4 min of the stimulus phase, restricting the use of a binomial model. Thus, we decided to score male responses only in the first 18 s after stimulus onset. This time corresponds to the median latency for any behavioural response of the focal males. To test the sensitiveness of our results to this criterion, we preformed the analysis using two other values: 12 s and 34 s, which correspond, respectively, to the upper limit of the first quartile and lower limit of the third quartile for any behavioural response of the focal males. As continuous predictor variables we used the background noise (in dB) and the light intensity (lux) at the signalling sites, both log-transformed, centred and standardized to zero (Table 2).

We performed all statistical analyses in the R version 3.6.1 (R Core Team, 2019).

TABLE 2 Summary of the variables and specifications of the generalized linear mixed models (GLMMs) performed to answer questions (1), (2), (3), (5), and (6) presented in Table 1. Question (4) is not included here because the analysis does not involve GLMM.

Questions	**Response variables**	**Type of response**	**Distribution family**	**Link function**	**Predictor variables**	**Type of predictor**	**Random variable**
(1), (2), and (3)	Aggressive notes, toe flags, other visual signals[a]	Continuous	Tweedie-Poisson[c]	Log	Control and stimulus phases	Categorical	Male identity
(5)	Aggressive notes, toe flags, other visual signals[b]	Continuous or binary	Gamma, binomial	Log (Gamma), cloglog[d] and logit (Binomial)	Male body condition and body size per stimulus phases	Continuous and categorical	Male identity
(6)	Response within 12, 18 and 34 s after stimulus onset[b]	Binary	Binomial	Cloglog[d] and logit	Background noise and light intensity per stimulus phases	Continuous and categorical	Male identity

[a] Emitted during all experimental phases of the visual, acoustic, and multimodal stimuli. [b] Emitted during only the stimulus phase of the visual, acoustic, and multimodal stimuli. [c] To account for underdispersion, overdispersion, and zero-inflation in count response variables (see Appendix S4). [d] To account for unbalanced number of zeros and ones in binary response variables (see Appendix S5, S7).

3 | RESULTS

Males responded to all stimuli types emitted by the EMR using aggressive notes, toe flags, and other visual signals, with a higher proportion of males responding to the acoustic and multimodal stimuli (Figure 2). Responses via aggressive notes and other visual signals were more frequent than via toe flags (Figure 2). These findings provide strong evidence of the efficacy of the EMR in eliciting male responses.

3.1 | Question (1): Is each signal (*a* and *v*) alone necessary or sufficient to elicit a receiver's response?

The stimulus phase of all experimental groups had an effect in the frequency of aggressive notes emitted by focal males (visual: $\beta \pm SE = 3.208 \pm 0.223$, $z = 14.40$, $p < 0.001$; acoustic: $\beta \pm SE = 4.582 \pm 0.152$, $z = 30.18$, $p < 0.001$; multimodal: $\beta \pm SE = 4.652 \pm 0.142$, $z = 32.86$, $p < 0.001$). However, only the stimulus phase of the acoustic and multimodal stimuli had an effect in the frequency of toe flags (acoustic: $\beta \pm SE = 1.450 \pm 0.305$, $z = 4.754$, $p < 0.001$; multimodal: $\beta \pm SE = 0.983 \pm 0.318$, $z = 3.090$, $p = 0.002$) and other visual signals emitted by focal males (acoustic: $\beta \pm SE = 0.443 \pm 0.202$, $z = 2.196$, $p = 0.028$; multimodal: $\beta \pm SE = 0.839 \pm 0.166$, $z = 5.073$, $p < 0.001$).

There were significant differences between experimental phases in the frequency of aggressive notes ($F_{8,288} = 25.59$, $p < 0.001$), toe flags ($F_{8,198} = 6.376$, $p < 0.001$), and other visual signals emitted by focal males ($F_{8,279} = 8.333$, $p < 0.001$). In short, the stimulus phases of the acoustic and multimodal stimuli elicited more aggressive notes, toe flags, and multimodal signals than their respective baseline pre-control phases, whilst the frequency of signals did not differ between the visual stimulus phase and the respective control phases. Thus, both the acoustic (*a*) and the multimodal stimuli (*a*+*v*) were sufficient to elicit a receiver's response, but the visual stimulus (*v*) was not (Figure 3).

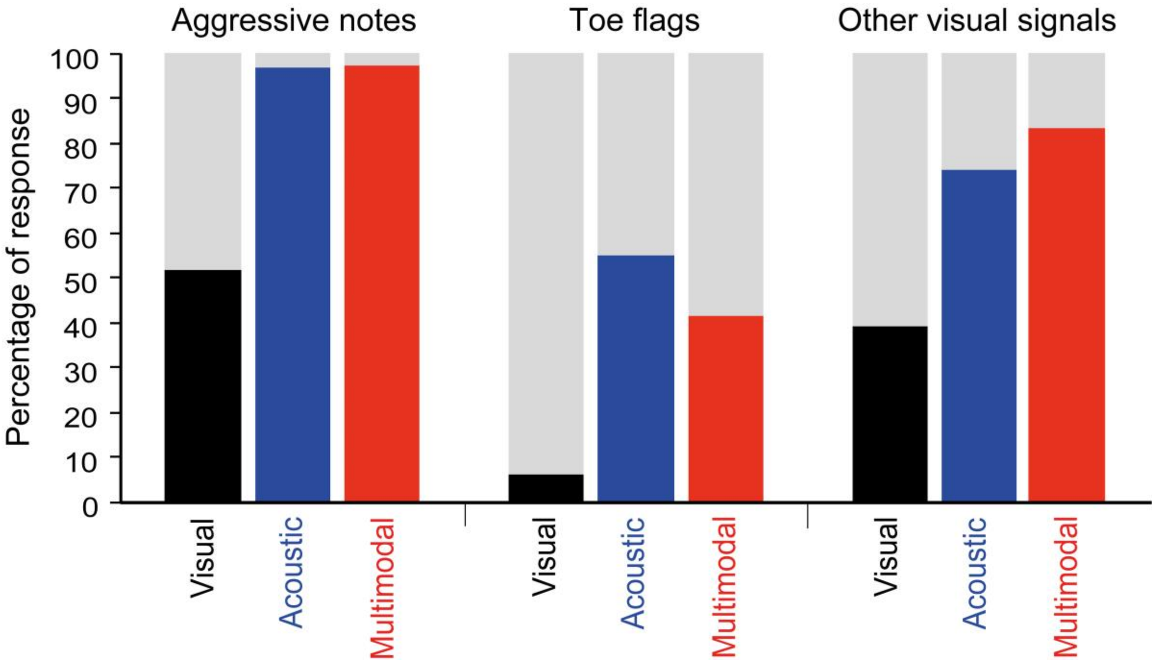

FIGURE 2 Percentage of *Crossodactylus schmidti* males that responded (black, blue, and red portions of the bars) with aggressive notes, toe flags, and other visual signals to visual (n = 33), acoustic (n = 31), and multimodal stimuli (n = 36) emitted by an electromechanical robot.

3.2 | Question (2): Are the expected receiver's responses to signals *a* and *v* redundant or non-redundant?

Because the acoustic stimulus was sufficient to elicit a response in the focal males, while the visual stimulus was not, the receiver's responses to *a* and *v* were necessarily non-redundant. When we compared only the stimulus phases between the visual and the acoustic experimental groups, the acoustic stimulus elicited more aggressive notes and toe flags than the visual stimulus. This finding indicates that the receiver's responses to *a* and *v* were quantitatively and qualitatively non-redundant (Figure 3).

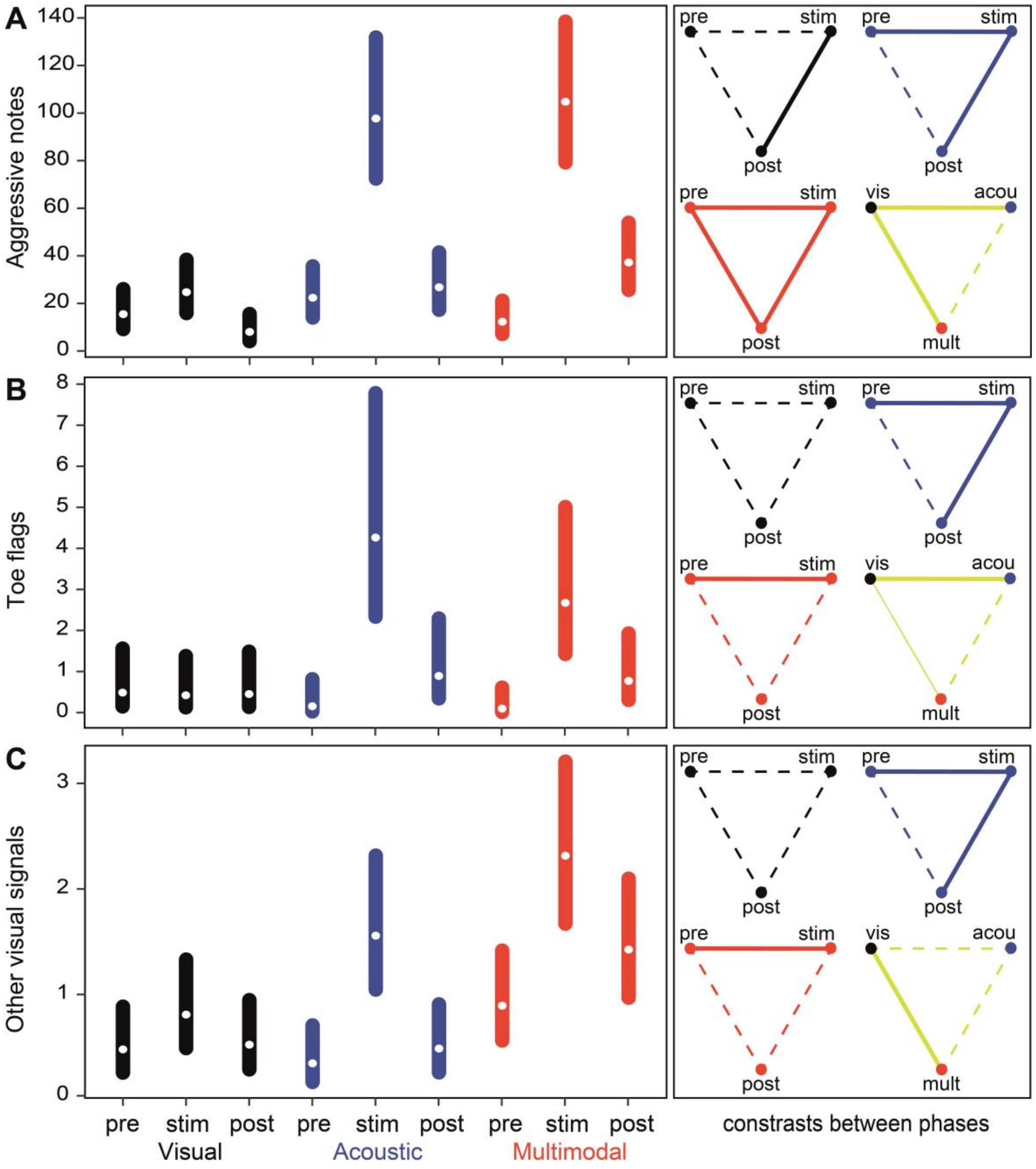

FIGURE 3 Aggressive notes (A), toe flags (B), and other visual signals (C) emitted by *Crossodactylus schmidti* males in response to visual, acoustic, and multimodal stimuli emitted by an electromechanical robot. White dots represent the mean response estimated by generalized linear mixed models for each experimental phase (*pre*-stimulus, stimulus, and *post*-stimulus), and the vertical bars represent 95% confidence intervals. Triangles in the right panels show the between-phase (black, blue, and red) and between-stimulus (yellow) contrasts. Thick solid lines connecting dots indicate significant *post-hoc* differences ($p < 0.05$); thin solid lines indicate marginally significant differences ($p = 0.05$); dashed lines indicate non-significant differences ($p > 0.05$). Detailed results of the GLMMs as well as the sample sizes of each model are presented in Appendix S4.

3.3 | Question (3): What is the category of the expected response to signal *a*+*v*?

In the post-control phase of the multimodal stimulus (*a*+*v*), males showed an increased response via aggressive notes compared to the pre-control phase (Figure 3A), which suggests modulation. Such an increase did not occur in response to *a* and *v* (Figure 3), and for this modulation effect to occur, *v* was necessary. Moreover, compared to the visual stimulus phase, the acoustic phase did not elicit a higher frequency of other visual signals (non-toe flags), but the multimodal stimulus phases did elicit, further suggesting a behavioural modulation in response to *a*+*v*.

3.4 | Question (4): Do the signals *a* and *v* covary in the receiver's response?

During the stimulus phase of the multimodal stimulus (*a*+*v*), the emission of toe flags (*v*) showed a positive correlation with the emission of aggressive notes (*a*) (r_s = 0.498, p = 0.011, Appendix S6). During the stimulus phase of the acoustic stimulus (*a*), however, the emission of toe flags did not correlate with the emission of aggressive notes (r_s = 0.183, p = 0.334, Appendix S6). The number of aggressive notes showed a positive correlation with the frequency of calls emitted both in response to *a* (r_s = 0.696, p < 0.001, Appendix S6) and *a*+*v* (r_s = 0.726, p < 0.001, Appendix S6). We did not test the correlation between signals for the visual stimulus (*v)* because only few responses via toe flags were recorded during the stimulus phase.

3.5 | Question (5): Do signals *a* and/or *v* covary with signaller quality?

During the stimulus phases, the frequency of aggressive notes (*a*) and the frequency of toe flags (*v*) emitted by the focal males was not influenced by body condition or size, irrespective of the type of stimulus (Figure 4, Appendix S5). Likewise, the frequency of other visual signals was not influenced by body condition or size (Figure 4, Appendix S5). Also, the probability of male's response via these different signals was not influenced by body condition or size (Appendix S5). For *v*, we did not test the influence of body condition and size on the emission of toe flags because few responses with toe flags were recorded during the stimulus phase.

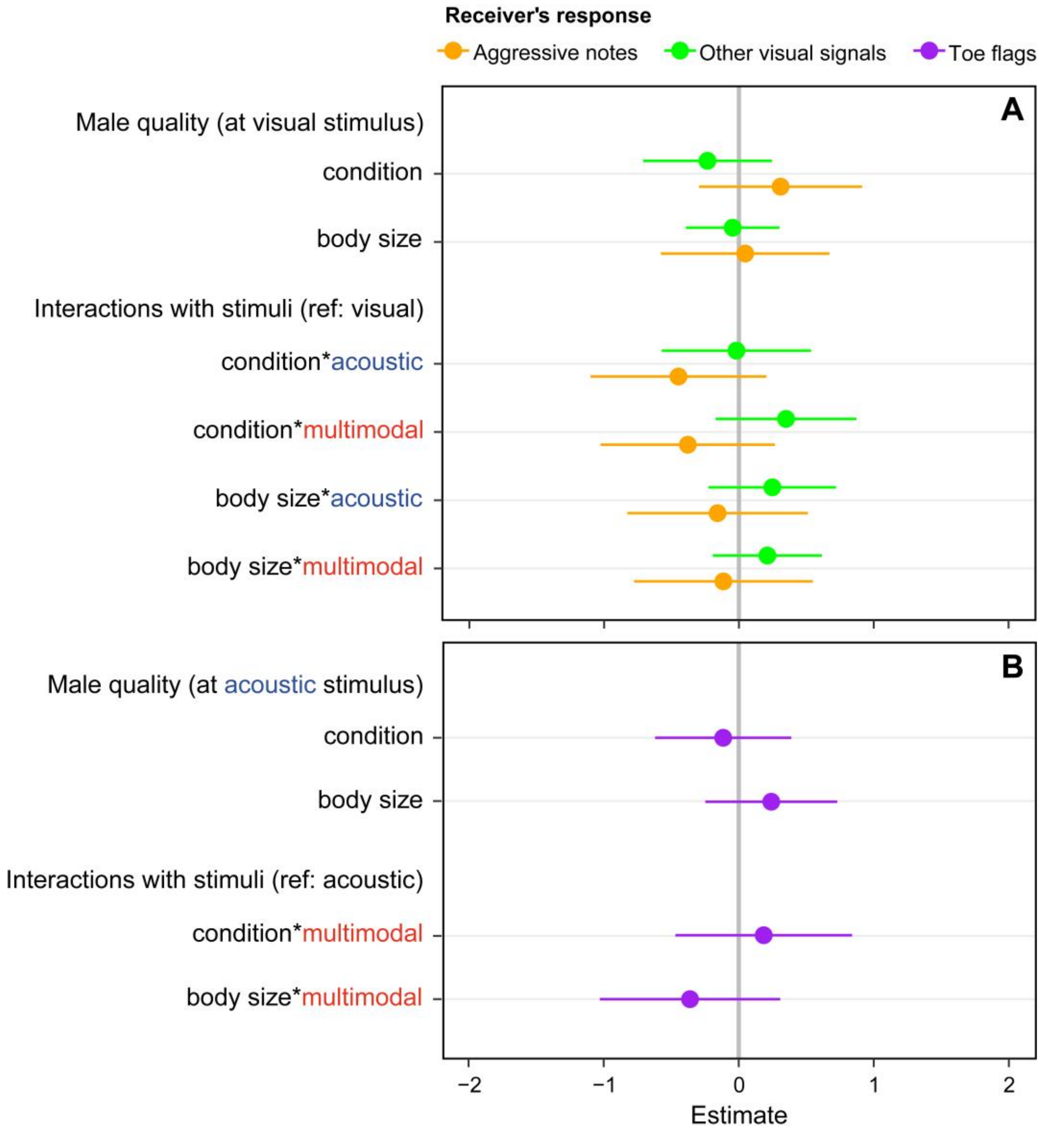

FIGURE 4 Results of the generalized linear mixed models (GLMMs) performed to investigate the effects of body condition and body size per type of stimulus (visual, acoustic, and multimodal) on the frequency of emission of aggressive notes, toe flags, and other visual signals by *Crossodactylus schmidti* males. Results are presented as mean estimate and standard error. Detailed results of the GLMMs as well as the sample sizes of each model are presented in Appendix S5.

3.6 | Question (6): Is the probability of response influenced by environmental conditions and higher in response to *a*+*v*?

Background noise and light intensity had no effect on the probability of receiver's response within 12 s or 18 s of visual (*v*), acoustic (*a*), or multimodal (*a*+*v*) stimuli onset. There was a marginally significant effect of light intensity on male late response within 34 s to *v* (Figure 5, Appendix S7). The probability of male response to *v* tended to increase with light intensity ($\beta \pm SE = 0.930 \pm 0.498$, $z = 1.869$, $p = 0.061$). The probability of male's response along the background noise and light intensity gradients was not higher in response to *a*+*v* than to *a* or *v* for any of the latency values.

4 | DISCUSSION

In this study we used an electromechanical robot (EMR) to test the function of a multimodal signal emitted by males of the frog *Crossodactylus schmidti*. Based on the knowledge on the ecology and behaviour of this species, we selected from the review by Hebets & Papaj (2005) five competing hypotheses contrasting three recognized types of selective pressures leading to the evolution of multimodal signals, from which we derived unique sets of predictions. The hypothesis that received most support was the *context* hypothesis (Table 1), which is an inter-signal interaction hypothesis that postulates that one signal provides a context in which a receiver can interpret and respond to a second signal. In what follows, we revise the questions of Table 1 based on our results and discuss why the context hypothesis provides the best explanation for the function of the multimodal signal emitted by *C. schmidti* males. We also explore the implications of our results to understand how one signal may provide a new context that alter a receiver's response to a second signal in animal contests.

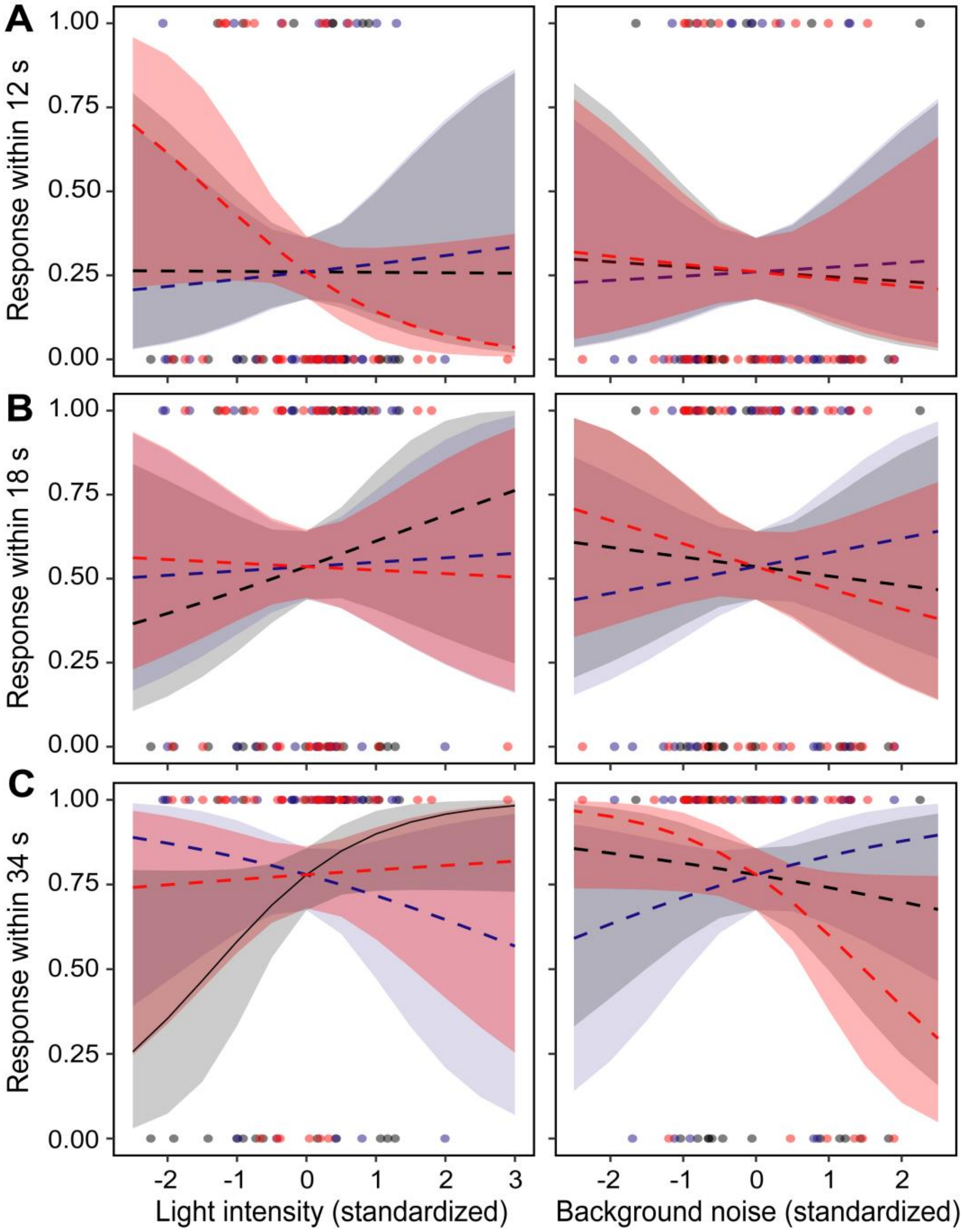

FIGURE 5 Effect of background noise and light intensity at the signalling sites of focal *Crossodactylus schmidti* males on their probability of response during the first (A) 12 s, (B) 18 s (C) or 34 s of visual (black dots and lines), acoustic (blue dots and lines), and multimodal stimulus (red dots and lines). Thin solid line indicates a marginally significant effect of light intensity (p = 0.06) in (C). Dashed lines in all graphics indicate non-significant effects (p > 0.06). Detailed results of the GLMMs as well as the sample sizes of each model are presented in Appendix S7.

In response to question (1), the unimodal acoustic signal (i.e., aggressive calls, signal *a*) was sufficient to elicit responses from male receivers in the form of aggressive notes, toe flags, and other visual signals — a variety of signals commonly observed during agonistic male-male encounters in *C. schmidti* (Caldart et al., 2014). The unimodal visual signal (i.e., toe flags, signal *v*), in turn, was not sufficient to elicit a response from male receivers, indicating that the signals *a* and *v* are non-redundant. From all hypotheses we tested, only the context hypothesis does not predict that each signal alone should elicit a receiver's response (Table 1). This hypothesis also predicts that the receiver's responses to each signal should be non-redundant, as we found here (see also question (2) in Table 1). Under natural conditions, toe flags are synchronized with aggressive calls (although the opposite is not true) and are rarely emitted unimodally (Caldart et al., 2014). This pattern makes sense if toe flags are added to aggressive calls to provide to the receiver a context for the interpretation of aggressive calls (see below).

Also in response to question (1), the multimodal stimulus (signal *a*+*v*) was sufficient to elicit receiver's responses, but not necessary because the acoustic stimulus alone elicited qualitatively and quantitatively similar agonistic responses in males during the stimulus phase. However, the multimodal stimulus elicited an acoustic response that lasted longer than the response to the unimodal stimulus, and also a higher response via other visual signals (non-toe flags) than the response to the visual stimulus. Thus, in response to question (3), our findings indicate that the addition of toe flags to aggressive calls modulates the receiver acoustic and visual response (Table 1), which implies that for the modulation effect to occur the emission of toe flags (signal *v*) is necessary. In a previous experiment that also used an EMR, males of the dart poison frog *Epipedobates femoralis* showed response patterns qualitatively similar to those described here (Narins et al., 2003). As in *C. schmidti*, the visual stimulus alone did not elicit aggressive responses from male receivers. Curiously, the acoustic stimulus alone rarely elicited agonistic responses from male receivers of *E. femoralis*, which engaged in physical contests against the EMR mostly when it emitted the multimodal stimulus. When the combined effect of non-redundant components modifies the receiver's responses compared to either component alone, we call it *modulation* (Partan & Marler, 2005). Although receiver's responses to the acoustic signal vary between *C. schmidti* and *E. femoralis*, in both species the multimodal stimulus clearly elicits modulation, which again is an exclusive prediction of the context hypothesis (Table 1).

One of the few empirical examples in favour of the context hypothesis has been reported for the snapping shrimp *Alpheus heterochaelis*, in which males are more responsive to the visual stimulus of an open claw in the presence of chemical signals from a male rather than a female (Hughes, 1996). This result suggests that chemical signals modify a male's response to the visual chela display, so that the multimodal signal has a context function: if the signaller is a male, the interaction may lead to aggression, but if the signaller is a female, the interaction may lead to pair-formation. Hughes (1996) argues that multimodal communication may be important if the same signal is used in different contexts, or by different classes of signallers, such as males and females, because receivers gain additional information to respond appropriately to different contexts. Aggressive calls are only emitted by males of *C. schmidti*, and thus our results provide a different example of a context function of a multimodal signal. From the signaller perspective, the emission of a multimodal signal composed of non-redundant components may be beneficial by making his aggressiveness or resource-defence potential (RHP) clearer to the receiver. Conversely, receivers exposed to toe flags simultaneously with aggressive calls may better interpret a rival intention of defending or invading a territory, which is important for decision-making in male-male contests (Dyson et al., 2013).

In response to question (4), the frequency of aggressive notes and toe flags emitted by males was positively correlated only in response to the multimodal stimulus, which indicates that the covariance between the acoustic and visual signals is not a fixed response to any type of signals emitted by rival males. If the multimodal signal has a context function, one could expect a positive correlation between signals in the receiver's responses to the multimodal signal emitted by another male. Simultaneous multimodal signals with a positive covariance between signals are predicted by the redundant, efficacy backup, and context hypotheses, though different underlying processes explain the expected correlation between signals (Hebets & Papaj, 2005). The redundant hypothesis predicts a correlation between signals because content-based selection drives multimodal signal evolution. The efficacy backup hypothesis predicts a correlation because efficacy-based selection favours signal integration due to environment pressures. Finally, the context hypothesis predicts a correlation because the between-signal integration itself is efficient in changing the receiver's response. During contests for territory possession, males of many anuran species may exchange signals conveying information on their RHP or even their residency status (Bee et al., 2016). This exchange of information may explain the correlation observed here between the receiver signals in response to a context multimodal signal of a potential rival, i.e., the EMR. To what extent this assortative form of communication mediated through a context multimodal signal indicates RHP or residency status remains unclear in the study species.

Regarding question (5), we found that the probability and frequency of aggressive notes, toe flags, and other visual signals emitted in response to the EMR did not correlate neither with body condition nor with male size, two proxies of male quality known to be important in male-male interactions of anurans (e.g., Băncilă et al., 2010; Dyson et al., 2013). Based on the content-based hypotheses that we have tested here (i.e., quality and redundant signal hypotheses), it is expected that only high-quality males would be able to pay the costs of high frequency of signal emission, resulting in a correlation between signal emission and male quality (Table 1). For the inter-signal interaction hypotheses (i.e., attention-altering and context hypotheses), in turn, the prediction is no correlation between signal emission and male quality (Table 1), as the key underlying process assumed to select the multimodal signal is the interaction between signals (Hebets & Papaj, 2005). Considering that the context hypothesis indeed provides the best functional explanation for the multimodal display of *C. schmidti*, we suggest that the simultaneous emission of toe flags and aggressive calls could indicate the level of motivation to fight using parameters that are arbitrary with respect to RHP (i.e., the so-called *conventional signals* sensu Guilford & Dawkins, 1995).

Finally, in response to question (6), there was no relationship between background noise or light intensity and the probability of receiver's response, with the exception of marginally significant effects of light intensity on male late response (34 s) to the visual stimulus. More importantly, we found that the probability of receiver's response to the multimodal stimulus was not higher along the noise or light gradient than the probability of response to the unimodal stimuli. These findings are consistent with predictions of content-based and inter-signal interaction hypotheses (Table 1). Among the hypotheses tested here, a higher probability of response to the multimodal stimulus is a prediction only of the efficacy backup hypothesis (Table 1), because species living in complex habitats may improve signal efficacy by adding to a hampered signal another one that is efficient in surpassing some transmission impediments (e.g., varying levels of noise or sunlight). Interestingly, a previous study showed that the daily calling activity of *C. schmidti* increases as light intensity increases (Caldart et al., 2016a). Another study showed that the advertisement calls are partially masked by the background noise of stream torrents (Caldart et al., 2016b). Thus, the above-mentioned marginal effect of light intensity on male probability of response warns that efficacy-based selection should not be ruled out as a possible explanation for other multimodal signals of *C. schmidti*. However, based on the lack of support to the efficacy backup hypothesis, we argue that efficacy-based selection alone does not explain the function of the multimodal display composed of aggressive calls and toe flags used in male-male agonistic interactions.

5 | CONCLUSION

The use of an electro-mechanical robot allowed us to experimentally test five competing hypotheses on the function of a multimodal signal composed of an acoustic and a visual component emitted simultaneously by males of the frog *C. schmidti*. The results, obtained entirely under field conditions, provide strong support to a context function of the multimodal display. We suggest that the addition of toe flags to aggressive calls modulates male-male interactions by providing to the receiver a new context for the interpretation and response to the acoustic signal. This new context may be motivation to fight or escalate a contest for territory possession, the level of aggressiveness, or intent of territory defence or invasion, depending on the identity of the signaller (i.e., resident or intruder). Accordingly, the context signal (toe flags) was necessary to magnify the receiver response, supporting the interpretation of its function in contest escalation. More broadly, our results indicate that inter-signal selection on multimodal signals is an important, though poorly investigated, driver of multimodal signal evolution in animals (see examples of an additional inter-signal interaction hypothesis in Grafe et al., 2007, 2012). Finally, our results also draw attention to the role of inter-signal selection on the evolution of multimodal signals in species living in heterogeneous habitats, where efficacy-based selection is often presumed to be a major driver (Rosenthal et al., 2004).

ACKNOWLEDGMENTS

We thank Tiago Theis, Álvaro Mainardi, Lívia Miron, Samanta Iop, Maiquel Elsenbach, and Maicon Elsenbach for helping in fieldwork, Sonia Cechin for lending field equipment, Solimary García-Hernández and André Rodrigues for assistance in statistical analyses, and Eileen Hebets for insightful discussions on our results.

FUNDING INFORMATION

VMC was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (2015/14336-6), and GM is supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (306550/2014-8).

AUTHOR'S CONTRIBUTIONS

VMC and GM conceived the study and designed methodology, VMC and MBS built the electromechanical robot, VMC collected the data, analyzed the data, and wrote a first version of the manuscript. VMC and GM led the writing of the manuscript. All authors revised the statistical procedures and the final version of the manuscript and gave final approval for publication.

DATA AVAILABILITY STATEMENT

Data and scripts will be available at Dryad Digital Repository upon manuscript acceptance.

ETHICAL NOTE

This project (ID 543) was authorized by the license number 21/2016 and the authorization number 173/2016 for temporary marking of individuals, both provided by the

Departamento de Biodiversidade, Divisão de Unidades de Conservação, Secretaria do Ambiente e Desenvolvimento Sustentável do Estado do Rio Grande do Sul (SEMA–RS), Brazil. No individual was collected or held in captivity during this study.

REFERENCES

Băncilă, R. I., Hartel, T., Plăiaşu, R., & Cogălniceanu, D. (2010). Comparing three body condition indices in amphibians: a case study of yellow-bellied toad *Bombina variegata*. *Amphibia-Reptilia*, *31*, 558-562. doi:10.1163/017353710X518405

Bee, M. A., Reichert, M. S., Tumulty, J. (2016). Assessment and recognition of competitive rivals in anuran amphibians. *Advances in the Study of Behavior, 48,* 161–249. doi:10.1016/bs.asb.2016.01.001

Betini, G. S., Avgar, T., & Fryxell, J. M. (2017). Why are we not evaluating multiple competing hypotheses in ecology and evolution? *Royal Society Open Science*, *4*, 160756. doi:10.1098/rsos.160756

Birkhead, T. R., Fletcher, F., & Pellat, E. J. (1998). Sexual selection in the zebra finch *Taeniopygia gutata*: condition, sex traits and immune capacity. *Behavioral Ecology and Sociobiology, 44*, 179–191. doi:10.1007/s002650050

Brooks, M. E., Kristensen, K., Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H.J., Maechler, M., Bolker, B.M. (2017). glmmTMB Balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal, 9*, 378-400. doi:10.32614/RJ-2017-066

Caldart, V. M., Iop, S., & Cechin, S. Z. (2011). Vocalizations of *Crossodactylus schmidti* Gallardo, 1961 (Anura, Hylodidae): advertisement call and aggressive call. *North-Western Journal of Zoology*, *7*, 118–124.

Caldart, V. M., Iop S., Bertaso T. R. N., & Cechin S. Z. (2012). Feeding ecology of *Crossodactylus schmidti* (Anura: Hylodidae) in Southern Brazil. *Zoological Studies*, *51*, 484–493.

Caldart, V. M., Iop, S., & Cechin, S. Z. 2014. Social interactions in a neotropical stream frog reveal a complex repertoire of visual signals and the use of multimodal communication. *Behaviour*, *151*, 719–739. doi:10.1163/1568539X-00003165

Caldart, V. M., Iop, S., Lingnau, R., & Cechin, S. Z. (2016a). Calling activity of a stream-breeding frog from the austral Neotropics: temporal patterns of activity and the role of environmental factors. *Herpetologica*, *72*, 90–97. doi:10.1655/HERPETOLOGICA-D-15-00029.

Caldart, V. M., Iop, S., Lingnau, R., & Cechin, S. Z. (2016b). Communication in a noisy environment: short-term acoustic adjustments and the underlying acoustic niche of a Neotropical stream-breeding frog. *Acta Ethologica*, *19*, 151–162. doi:10.1007/s10211-016-0235-2

Caldart, V. M., Loebens, L., Brum, A. J. C., Bataioli, L., & Cechin, S. Z. (2019). Reproductive cycle, size and age at sexual maturity, and sexual dimorphism in the stream-breeding frog *Crossodactylus schmidti* (Hylodidae). *South American Journal of Herpetology*, *14*, 1–11. doi:10.2994/SAJH-D-17-00060.1.

Dyson, M. L., Reichert, M. S., & Halliday, T. R. (2013). Contests in amphibians. In I. C. W Hardy, & M. Briffa (Eds.), *Animal contests* (pp. 228-257). Cambridge, UK: Cambridge University Press.

Endler, J. A. (1992). Signals, signal conditions, and the direction of evolution. *The American Naturalist*, 139, S125–S153. doi:10.1086/285308

Freeman, A. R., & Hare, J. F. (2015). Infrasound in mating displays: a peacock's tale. *Animal Behaviour*, *102*, 241–250. doi:10.1016/j.anbehav.2015.01.029

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., & Hothorn, T. (2020). mvtnorm: Multivariate Normal and t Distributions. R package version 1.1-0. Available at: https://CRAN.R-project.org/package=mvtnorm.

Guilford, T., & Dawkins, M. S. (1995) What are conventional signals? *Animal Behaviour*, *49*, 1689-1695. doi: 0.1016/0003-3472(95)90090-X

Grafe, T. U., & Wanger, T. C. (2007). Multimodal signaling in male and female foot-flagging frogs *Staurois guttatus* (Ranidae): an alerting function of calling. *Ethology*, *113*, 772–781. doi:10.1111/j.1439-0310.2007.01378.x

Grafe, T. U., Preininger, D., Sztatecsny, M., Kasah, R., Dehling, J. M., Proksch, S., & Hödl, W. 2012. Multimodal communication in a noisy environment: a case study of the Bornean rock frog *Staurois parvus*. *PLoS One*, 7:e37965. doi:10.1371/journal.pone.0037965

Hartig, F. (2020). DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. R package version 0.2.7. Available at: https://CRAN.R-project.org/package=DHARMa.

Hartmann, M. T., Giasson, L. O. M., Hartmann, P. A., & Haddad, C. F. B. (2005). Visual communication in Brazilian species of anurans from the Atlantic forest. *Journal of Natural History*, *39*, 1675–1685. doi:10.1080/00222930400008744

Hebets, E. A., & Papaj, D. R. (2005). Complex signal function: developing a framework of testable hypotheses. *Behavioral Ecology and Sociobiology, 57*, 197–214. doi:10.1007/s00265-004-0865-7

Higham, J.P. & Hebets, E. A. (2013). An introduction to multimodal communication. *Behavioral Ecology and Sociobiology*, *67*, 1381–1388. doi:10.1007/s00265-013-1590-x

Hödl, W., & Amézquita, A. (2001). Visual signalling in anuran amphibians. In M. J. Ryan (Ed.), *Anuran communication* (pp. 121–141). Washington, DC: Smithsonian Institution Press.

Hölldobler, B. (1999). Multimodal signals in ant communication. *Journal of Comparative Physiology A*, *184*, 129–141. doi:10.1007/s003590050313.

Hughes, M. (1996). The function of concurrent signals: visual and chemical communication in snapping shrimp. *Animal Behavior 52*: 247–257. doi:10.1006/anbe.1996.0170

Lenth, R. (2019). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.4. Available at: https://CRAN.R-project.org/package=emmeans.

Maluf, J. R. T. (2000). Nova classificação climática do Estado do Rio Grande do Sul. *Revista Brasileira de Agrometeorologia*, *8*, 141–150.

Narins, P. M., Hödl, W., & Grabul, D. S. (2003). Bimodal signal requisite for agonistic behavior in a dart-poison frog, *Epipedobates femoralis*. *Proceedings of the National Academy of Science, United States of America*, *100*, 577–580. doi:10.1073/pnas.0237165100

Partan, S., & Marler, P. (2005). Issues in the classification of multimodal communication signals. *The American Naturalist*, *166*, 231–245. doi:10.1086/431246

R Core Team. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing. v.3.6.1. Foundation for Statistical Computing, Vienna, Austria.

Rosenthal, G. G., Rand, A. S., & Ryan, M. J. (2004). The vocal sac as a visual cue in anuran communication: an experimental analysis using video playback. *Animal Behaviour*, *68*, 55–58. doi:10.1016/j.anbehav.2003.07.013

Starnberger, I., Preininger, D., & Hödl, W. 2014. From uni- to multimodality: towards an integrative view on anuran communication. *Journal of Comparative Physiology A*, *200*, 777–787. doi:10.1007/s00359-014-0923-1

Wilgers, D. J., & Hebets, E. A. (2012). Seismic signaling is crucial for female mate choice in a multimodal signaling wolf spider. *Ethology*, *118*, 387–397. doi:10.1111/j.1439-0310.2012.02023.x

Zahavi, A. (1975). Mate selection—a selection for a handicap. *Journal of Theoretical Biology*, 53, 205-214. doi:10.1016/0022-5193(75)90111-3