Journal of the
Experimental Analysis of Behavior
JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 2014, **101,** 215–229 NUMBER 2 (MARCH)

RESPONDING BY EXCLUSION IN TEMPORAL DISCRIMINATION TASKS

Nathália Sabaine Cippola[1], Camila Domeniconi[1], and Armando Machado[2]

[1]FEDERAL UNIVERSITY OF SÃO CARLOS, BRAZIL
[2]UNIVERSITY OF MINHO, PORTUGAL

Responding by exclusion, one of the most robust phenomena in Experimental Psychology, describes a particular form of responding observed in symbolic, matching-to-sample tasks. Given two comparison stimuli, one experimentally defined and one experimentally undefined, the participant prefers the undefined comparison following an undefined sample. The goal of the present study was to determine whether responding by exclusion could be obtained using samples that varied along a single dimension. Using a double temporal bisection task, 10 university students learned to choose visual comparisons (colored circles) based on the duration of a tone. In tests of exclusion, sample stimuli with new durations were followed by comparison sets that included one previously trained, defined comparison (colored circle) and one previously untrained, undefined comparison (geometric shape). Participants preferred the defined comparisons following the defined samples and the undefined comparisons following the undefined samples, the choice pattern typical of responding by exclusion. The use of samples varying along a single dimension allows us to study the interaction between stimulus generalization gradients and exclusion in the control of conditional responding.

Key words: responding by exclusion, temporal stimuli, stimulus control, conditional relation, university students

---

The term "responding by exclusion" describes a particular form of responding observed in symbolic, matching-to-sample tasks. Suppose that, when exposed to comparison stimuli $C_1$and C_2 after the presentation of the sample stimulus S_1or S_2, a participant learns to choose C_1 after S_1 and C_2 after S_2. We represent these conditional discriminations as $S_1 \rightarrow C_1$ and $S_2 \rightarrow C_2$. Subsequently, after the presentation of a sample stimulus S_X, a sample not previously related to any comparison stimulus, the participant is given a choice between C_1 and C_Y, or C_2 and C_Y, in which the comparison C_Y is unrelated to any previous sample. The sample S_X and the comparison C_Y are both undefined stimuli. These relationships can be represented as follows

$$\text{Training}: \begin{cases} S_1 \rightarrow C_1 \\ S_2 \rightarrow C_2 \end{cases}$$
$$\text{Testing}: \begin{cases} S_X \rightarrow C_Y \text{ or } C_1 \\ S_X \rightarrow C_Y \text{ or } C_2 \end{cases}.$$

Several studies have shown that, given the undefined sample S_X, the participant tends to exclude C_1 or C_2, the comparisons previously related to samples other than the current sample, and choose C_Y, the comparison that also is undefined. The preference for C_Y given S_x occurs without any explicit training or teaching of the $S_X \rightarrow C_Y$ relation. In addition to responding by exclusion, the expression used by behavioral analysts (e.g. Costa, Wilkinson, McIlvane, & de Souza, 2001; Dixon, 1977; McIlvane, Kledaras, Lowry & Stoddard,1992; Stromer, & Osborne, 1982; Wilkinson, Rosenquist, & McIlvane, 2009), this response pattern has been termed "fast mapping" by psycholinguists (e.g., Bates, Benigni, Bretherton, Camaioni, & Volterra, 1979; Carey & Bartlett, 1978; Floor & Akhtar, 2006; Halberda, 2003;

The first author was supported by a master's degree fellowship by the Ministry of Education (CAPES). Armando Machado was supported by grant PTDC/MHC-PCN/3540/2012 from the Portuguese Foundation for Science and Technology. Camila Domeniconi had a post-doctoral fellowship from the Foundation for Research Support in the State of São Paulo (FAPESP, 2009/18479-5). She is currently affiliated with the National Institute of Science and Technology on Behavior, Cognition and Teaching. Grants: FAPESP (08/57705-8) and CNPq (573972/2008-7). She has a research productivity fellowship by the National Council for Scientific and Technological Development (CNPq, 301623/2012-0).

Address correspondence to Camila Domeniconi, Programa de Pós-Graduação em Psicologia, Universidade Federal de São Carlos, Rod. Washington Luís, km 253. Caixa Postal 676, 13565-905, São Carlos, SP (email: domeniconicamila@gmail.com).

doi: 10.1002/jeab.71

Pruden, Hirsh-Pasek, Golinkoff, & Hennon, 2006).

Responding by exclusion was first observed in a study on the teaching of conditional discriminations to adolescents with intellectual disability (Dixon, 1977). To study the conditional control that a verbal stimulus exerts on the participant's choice, Dixon programmed training trials that included Greek letters or Japanese characters as comparisons and the corresponding dictated words as samples. To illustrate, on some trials, Dixon dictated the word "pi" and then presented two visual comparisons, the Greek letter π, the correct choice or S^+, and another letter that could be either θ or Y, the incorrect choice or S^-. Once the relation "pi"→π was learned, exclusion tests were performed. A new, undefined dictated word such as "theta" was presented as sample followed by two visual comparisons, π and θ. The participants systematically chose the undefined visual stimulus θ after the new sample "theta", but continued to choose the defined visual stimulus π after the defined sample "pi". Similarly, following the undefined sample "epsilon", the participants chose the undefined comparison Y over the defined comparison π, but continued to choose π following the defined sample "pi". Based on these findings, Dixon suggested that the participants responded to the new comparisons "by exclusion": When the dictated word was unknown or undefined, the participants excluded the defined or trained visual stimulus and chose the undefined, untrained visual stimulus.

Since then several studies have attested to the robustness of responding by exclusion in participants with typical and atypical development (Dixon, Dixon, & Spradlin, 1983; McIlvane et al., 1992; McIlvane, Munson, & Stoddard, 1988; McIlvane & Stoddard, 1981; Stromer, 1986, 1989), in children of different ages (Costa et al., 2001; Domeniconi, Costa, de Souza, & de Rose, 2007; Ferrari, de Rose, & McIlvane, 1993), and in university students (McIlvane, Kledaras, Munson, King, de Rose, & Stoddard, 1987).

Other studies have suggested that some animals also respond by exclusion, including sea lions and dolphins (Herman, Richards, & Wolz, 1984; Kastak & Schusterman, 2002; Schusterman & Krieger, 1984), chimpanzees (Beran & Washburn, 2002) and dogs (Aust, Range, Steurer, & Huber, 2008; Kaminski, Call, & Fischer, 2004; Pilley and Reid, 2011). Kaminski et al. (2004) investigated exclusion performance in a border collie in a task in which the dog should retrieve objects after a proper-noun verbal command. The available objects were seven familiar objects and one new object. The dog's task consisted of going into an experimental room and retrieving the object that corresponded to the name spoken by the experimenter. On the first trial, the experimenter always requested a familiar object; on the second or third trials, the experimenter requested the new object. In 7 out of 10 trials, the dog excluded the familiar objects and chose the unfamiliar one. The authors argued that the dog related the unfamiliar word and the unfamiliar object without explicit training, thereby showing learning by exclusion of a new auditory–visual relation.

Other comparative studies investigated exclusion performance in animals using alternative procedures to conditional discrimination such as simple discriminations and food-finding tasks (Aust et al., 2008; Beran, 2010; Call, 2006; Erdöhegyl, Topál, Virányl, & Miklósi, 2007; Schloegl, Bugnyar, & Aust, 2009). Taken together, the results of previous studies, particularly with humans, have led some authors to argue that the pattern of responding by exclusion is a very robust effect in experimental psychology (Wilkinson, de Souza, & McIlvane, 2000).

However, researchers still do not know what are the minimal conditions required to obtain responding by exclusion. For example, Pruden et al. (2006) investigated the mapping of words onto objects by 10-month-old infants and found that infants were more likely to respond by exclusion when the test trials included perceptually salient objects. The authors concluded that the use of perceptually salient objects is important during the beginning of the word-learning process.

Horst, Samuelson, Kucker and McMurray, (2011) manipulated the novelty of the comparison stimulus by exposing two-year-old children to the novel object before the exclusion tests. During the tests, the children were presented with three novel objects as comparison stimuli, two to which they were pre-exposed and one completely new, and a new name as the sample stimulus. The children chose the completely new stimulus with highest probability. The authors concluded that endogenous biases to

novelty play an important role when children map novel names onto objects.

Costa, de Rose and de Souza, (2010) investigated how responding by exclusion changes when verbal, auditory samples included lexical cues during testing. These lexical cues could denote plurals (e.g., "mopadis" instead of "mopadi"), size (e.g.,"mopadinho", with the Portuguese diminutive "inho" meaning "small"), or action (e.g., "mopadiando", similar to the English "ing" form). The results showed that adding the lexical cues during the test trials reduced the choices of the undefined comparison stimulus, that is, of responding by exclusion.

The foregoing studies revolve around the notion of an undefined stimulus and point to the need to clarify that notion. If one thinks that any stimulus can be characterized by the specific values it has along its various dimensions, regardless of how complex these dimensions might be, then the notion of an undefined stimulus can be seen in a new light, as a matter of degree rather than all-or-none. Fields, Matneja, Varelas, Belanich, Fitzer, & Shamoun (2002) investigated the perceptual classes as a set of stimuli arrayed along some continuum and discussed the adaptive advantage of perceptual classes in stimulus control because they enable the organism to respond appropriately to undefined stimuli without direct training. Sample or comparison, a stimulus will be more undefined the more its values along the relevant dimensions differ from the corresponding values of the defined stimuli. Or, to put it in a different but equivalent way, the multidimensional generalization gradients of the defined stimuli set the criteria to define how "undefined" a new stimulus is. The goal of the present study was to explore this conception and its implications for responding by exclusion.

To that end, we simplified the stimulus samples so that they varied along a single dimension, duration. The defined samples corresponded to tones with distinct durations (e.g., 200 ms and 600 ms) and they were trained in the conditional relations $S_{200} \rightarrow C_1$ and $S_{600} \rightarrow C_2$. The undefined samples S_X corresponded to tones with durations that (a) were not previously trained in any conditional relation and (b) were located at different distances from the trained values of 200 and 600ms. We expected that this new procedure would allow us to observe how the sample controlled the participant's responding in general and his/her responding by exclusion in particular.

We predicted that, given a choice between an undefined comparison (C_Y) and a defined comparison (C_1 or C_2), responding by exclusion would be more likely as the duration of the undefined sample S_X moved away from the duration of the defined sample associated with C_1 or C_2 (i.e., S_{200} if C_1 and S_{600} if C_2).In other words, the probability of choosing the undefined comparison C_Y should increase with the difference between the test sample and the sample related to the alternative comparison.

The preceding argument is incomplete because it fails to distinguish two seemingly different processes, responding by exclusion and generalization decrement. Having learned to choose C_2 following S_{600} samples (and perhaps not to choose C_2 following S_{200} samples), the temporal generalization gradient associated with C_2 would probably have a minimum close to 200 ms and a maximum close to 600 ms (e.g., Wearden, 1991; Wearden & Lejeune, 2008). Hence, due to temporal generalization, the tendency to choose C_2 over C_Y should increase (and the tendency to choose C_Y over C_2 should decrease) as the sample duration ranges from 200 to 600 ms. Thus, to explain how preference for C_Y varies with sample duration, temporal generalization suffices; responding by exclusion is not necessary. However, a generalization-based account would not predict a strong preference for C_Y over C_2 following samples significantly longer than 600 ms (e.g., 5-s samples). Such samples were not associated either positively or negatively with the two comparisons and therefore if a strong preference for the undefined comparison is observed, then it may provide evidence for responding by exclusion. A similar argument applies for samples significantly shorter than 600 ms: At best, generalization decrement predicts indifference between C_Y and C_2, whereas responding by exclusion predicts a clear preference for C_Y. In summary, we hypothesize that, following samples outside the basin of the generalization gradient, a generalization-based account predicts at best indifference between C_2 and C_Y, whereas a responding-by-exclusion account predicts a preference for C_Y clearly above .5.

Method

Participants

Ten university students, 18 to 26 years of age, participated in the experiment. They were naive with regard to experimental procedures related to temporal discrimination and exclusion. Data collection began after the ethics committee of the Federal University of São Carlos approved the study (protocol CAAE 0022.0.135.000-10) and the participants signed an informed consent form.

Experimental Settings and Materials

The experimental sessions were conducted in a quiet room of the Laboratory of Studies of Human Behavior at the Federal University of São Carlos. The acoustically isolated room was equipped with one table, two chairs, a portable computer running the ProgMTS software (Marcicano, Carmo, & Prado, 2011), a mouse, two speakers, and a video camera.

Stimuli

The experimental task consisted of using the computer mouse to choose visual stimuli on the computer screen after the presentation of the acoustic sample. The sample stimuli were 500-Hz, 70-dB tones that varied in duration. The tones were presented through the speakers connected to the computer. The defined comparison stimuli were four circles with different colors: red (R), green (G), blue (B), and yellow (Y). The undefined comparison stimuli were six geometric, black-and-white shapes, henceforth designated by U_1, U_2, …, and U_6. Geometric shapes were used to reduce the likelihood of stimulus generalization from the colored circles. All comparison stimuli were 3.0 cm wide by 3.0 cm tall.

General Procedure

The participants first learned two temporal discriminations. In the first discrimination, the participant heard a 200- or 600-ms tone at the beginning of each trial. Then, two circles, R and G, appeared in a random location on the computer screen. The choice of R was correct after the 200-ms tone and the choice of G was correct after the 600-ms tone (i.e., S_{200}→R and S_{600}→G). In the second discrimination, the procedure was similar, except that the participant heard a 600- or 1800-ms tone and then chose between two other circles, B and Y, with the correct mapping being S_{600}→Band S_{1800}→Y. Importantly, the longer sample in the first discrimination was the shortest sample in the second discrimination. This task has been used with pigeons to study the effect of the learning context on temporal discriminations (see Machado, Malheiro, & Erlhagen, 2009; Machado & Pata, 2005; Oliveira & Machado, 2007), the interaction among temporal generalization gradients (Vieira de Castro & Machado, 2010, 2012), and the emergence of new conditional relations (Huziwara, Velasco, Tomanari, de Souza, & Machado, 2012).

The study was divided into four training phases and one testing phase. Table 1 shows the structure of each training phase. To facilitate learning and reduce the number of errors, during Phases 1 and 2 we introduced the comparison stimuli gradually and blocked the trials (e.g., Saunders & Spradlin, 1990). Specifically, in Phase 1, during the first four trials only the S_{200} stimulus was presented as a sample and only the R circle (i.e., the correct response) was presented as the comparison. On the next eight trials, only the S_{200} stimulus continued to be presented as a sample, but now two comparison circles appeared. On the next four trials, only

Table 1

Number of trials, sample and comparison stimuli presented in each training phase.

Phase	Trials	Sample	Comparisons
1	4	S_{200}	R
	8	S_{200}	R + G
	4	S_{600}	G
	8	S_{600}	R + G
	36	S_{200} or S_{600}	R + G
2	4	S_{600}	B
	8	S_{600}	B + Y
	4	S_{1800}	Y
	8	S_{1800}	B + Y
	36	S_{600} or S_{1800}	B + Y
3	8	S_{200}	R + G
	8	S_{600}	R + G
	8	S_{600}	B + Y
	8	S_{1800}	B + Y
4	9 + 12	S_{200}	R + G
	9 + 12	S_{600}	R + G
	9 + 12	S_{600}	B + Y
	9 + 12	S_{1800}	B + Y

Note. In Phase 4, the two trial numbers indicate the number of reinforceable (9) and nonreinforced (12) trials. Sample duration is in milliseconds (e.g., M_{200} = 200 ms). The comparison stimuli were circles of different colors (R = red, G = green, B = blue, and Y = yellow).

the S_{600} sample was presented, and only the G circle (the correct response) appeared on the computer screen. For the next eight trials, the two circles appeared. Lastly, on the remaining 36 trials, the two samples were presented 18 times each, in a random order, followed by the two comparisons, Rand G. Each correct response was followed by an oral expression of social approval presented by the computer (e.g., "Very good," "Great," "OK," "Excellent," or "Congratulations"). Incorrect responses did not have any programmed consequences. A 1-s interval separated consecutive trials.

At the beginning of Phase 1, the experimenter gave the participants the following instructions: "You will perform a task in which you have to use the mouse to click on a circle or on another stimulus when they appear on the screen after the beep."

In Phase 2, all of the procedural details remained the same as in Phase 1, except that the samples and comparisons belonged to the second discrimination ($M_{600}\rightarrow$ Band $S_{1800}\rightarrow$ Y). In Phase 3, the four types of trials occurred in the same session in a random order. In addition, a criterion was introduced to advance to the next phase. If, for any of the four trial types, the participant made more incorrect than correct responses, then Phase 3 was repeated.

The goal of Phase 4 was to adapt the participants to the absence of reinforcement that would be present during the subsequent test trials. To this end, the participants received the following instructions: "You will have to perform the same task as before. However, not all of your responses will be praised." For each sample, there were 9 reinforceable trials (in case of correct choice) and 12 unreinforced trials (even if choice was correct). Phase 4 continued until the participants made at most one error in each bisection task. Each participant required from 1 to 3 training sessions. Each session lasted 15 min, occurred generally once per day, and was video recorded.

Phase 4 was followed by the test phase that determined how the participants chose between a defined comparison stimulus (R, G, B, or Y) and an undefined comparison stimulus (U_1 to U_6) conditional on sample stimuli with different durations. The instructions were as follows: "You will continue to perform the same task of clicking one of the figures that appears on the screen after the beep."

Table 2 shows the structure of each session. The trials were of two sorts, baseline trials and test trials. The baseline trials were similar to the previous training trials and allowed us to monitor the accuracy of the base discriminations. The test trials included six samples varying in duration, three with the previously trained durations (i.e., 200, 600, and 1800 ms) and three with new durations (i.e., 350, 1040, and 7200 ms). Of the new durations, the first two corresponded to the geometric means of the pairs of trained durations (350 is approximately the geometric mean of 200 and 600, and 1040 is approximately the geometric mean of 600 and 1800); the 7200-ms test duration was substantially greater than the longest duration used during training. These durations were chosen because they cover a relatively large range of values and are all clearly discriminable by most human participants—each successive duration in the series is at least 73% longer than its predecessor, a value significantly greater than the typical Weber fraction of 20% obtained in human temporal bisection tasks (see Wearden,1991; Wearden & Lejeune, 2008).

Table 2

Number of trials, sample and comparison stimuli presented in each session of the testing phase.

Baseline			Exclusion Test		
Trials	Samples	Comparisons	Trials	Samples	Comparisons
9	S_{200}	R + G	4	S_{200}	$C_i + U_1$
9	S_{600}	R + G	12	S_{350}	$C_i + U_2$
9	S_{600}	B + Y	4	S_{600}	$C_i + U_3$
9	S_{1800}	B + Y	12	S_{1040}	$C_i + U_4$
			4	S_{1800}	$C_i + U_5$
			12	S_{7200}	$C_i + U_6$

Note. On the exclusion test trials, $U_1, U_2, \ldots, U_6$ refer to the six undefined comparisons (black and white geometric shapes) and C_i, with U = R, G, B, or Y, refers to the four defined stimuli (circles of different colors). The frequency of each C_i was equal to the number of trials divided by four.

Each test sample was followed by two comparisons, an undefined comparison (shape) that remained the same for each sample, and one of the four defined comparisons (colors; see last column of Table 2). Each of the three previously trained durations was presented four times, one for each defined comparison, and each of the three new durations was presented 12 times, four times for each defined comparison. All test trials were performed in extinction, even those that used previously trained samples. In each testing session, the baseline and test trials occurred in random order. Five testing sessions were conducted, generally one session per day; each session lasted approximately 15 min.

At the end of the study, in a brief interview, the participants responded to three questions: (1) "Was there any criterion to click on one or the other circle? If yes, what was it?" (2) "What was your choice criterion when there was a colored circle and a figure that you had not seen before?" and (3) "Did you count to measure time? If so, how did you do it?" The responses were recorded on video and analyzed subsequently. Finally, as a reward for participating in the study, the students received a coupon exchangeable for photocopies at the university store; the value of the coupon was the same for all participants.

Results

The average percentage of correct responses in training Phases 1, 2, and 3 was 89%. In Phase 4, with intermittent reinforcement of correct responses, the participants needed an average of 2.3 sessions to meet the criterion, and the average percentage of correct responses equaled 97%. During the five testing sessions, correct responses on baseline trials equaled 80%. Although lower than in the training phases, the accuracy of the base discriminations remained high.

Figure 1 shows the average performance on the exclusion tests. Consider the curve that corresponds to the R comparison (filled circles). Consistent with our hypothesis, the choice of the undefined stimulus was more likely as the duration of the sample moved away from 200 ms, the duration associated with R during training. Critically, the proportion of choices of the undefined stimulus increased to 1 for longer samples.

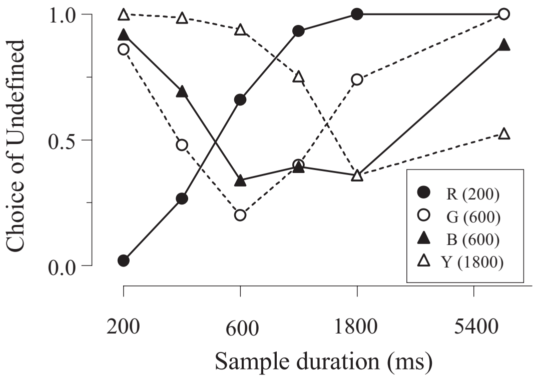

Fig. 1. The curves show the average proportion of choices of the undefined stimulus as a function of sample duration with the defined comparison stimulus (R, G, B and Y) as a parameter. The values in parenthesis show the sample durations associated during training with the respective comparisons. The x-axis is on a log scale.

The results for the G (empty circles) and B (filled triangles) comparisons, both associated with the 600-ms sample during training, revealed the same effect: The preference for the undefined stimulus was more likely as the duration of the sample moved away from 600 ms. Variability across participants explains why the choice proportions at 600 ms, the sample associated with G and B, was not lower. We analyze individual performance below.

The results for the Y comparison (empty triangles) also revealed an increasing preference for the undefined stimulus as the duration of the sample moved away from 1800 ms, although this increase was only substantial for shorter samples.

In summary, the results of the four tests were consistent with the hypothesis that the more the test sample differed from the sample paired during training with the comparison present in the choice set (i.e., the more undefined the sample was with respect to the defined comparison stimulus), the more likely was the choice of the undefined comparison.

To evaluate the statistical significance of the results, we performed four analyses of variance (ANOVAs), one for each defined comparison stimulus, with duration as the repeated measure. All of the analyses revealed a strong effect of the duration of the sample on the preference for the undefined stimulus ($F_{5,45} > 7.7$, $p < .01$). In this context, it is of particular interest to compare the two curves for the 600-ms sample. Visual

inspection of Figure 1 suggests that the curves are similar in shape but not in location because the B curve is shifted to the right compared with the G curve. A repeated-measures ANOVA with sample duration (six levels) and comparison stimulus (two levels) as factors revealed a significant interaction ($F_{5,45} = 5.34$, $p < .01$). As described below, closer analysis of the G and B curves revealed two distinct groups of participants, one that produced very similar, overlapping G and B curves, and another that produced a B curve to the right of the G curve.

To better understand the nature of individual differences, Figures 2 and 3 show the data from all participants. Figure 2 shows the results of the four participants that produced overlapping G and B curves. In the tests with these comparisons, the proportion of choices of the undefined stimulus followed a U-shaped curve, with a minimum at or close to 600 ms. For all participants, the choice proportion at 600 ms was significantly below 0.5. In the tests with the R comparison, the four participants displayed a systematically increasing curve. In the tests with the Y comparison, the differences among participants occurred to a small extent at the 1800-ms sample and to a larger extent at the 7200-ms sample.

The results of S10 are noteworthy because they are the closest to our hypothesis. In each curve, the minimum occurs at the defined sample for that curve (i.e., 200 ms for R, 600 ms for G and B, and 1800 ms for Y). In addition, as sample duration moves away from the training value, choice proportion increases, not to indifference, as temporal generalization predicts, but to 1, as responding by exclusion predicts.

Figure 3 shows the results from the participants that produced nonoverlapping G and B curves. Five of the six participants (except S3) produced a B curve to the right of the G curve. For the G curve, the minimum occurred at or close to 600 ms for five participants and choice proportion at the minimum ranged from .5 to 0. For the B curve, the minimum occurred at 1800 ms for four participants and in all cases choice proportion at the minimum equaled 0. For the remaining two participants, the minimum occurred at 350 ms (S3) and at durations equal to or greater than 600 ms (S7). In the tests with the R comparison, all participants showed an increasing curve, with a minimum at 200 ms. Finally, in the tests with the Y comparison, for all participants choice proportion equaled or was close to 1 following the shortest samples. Then,

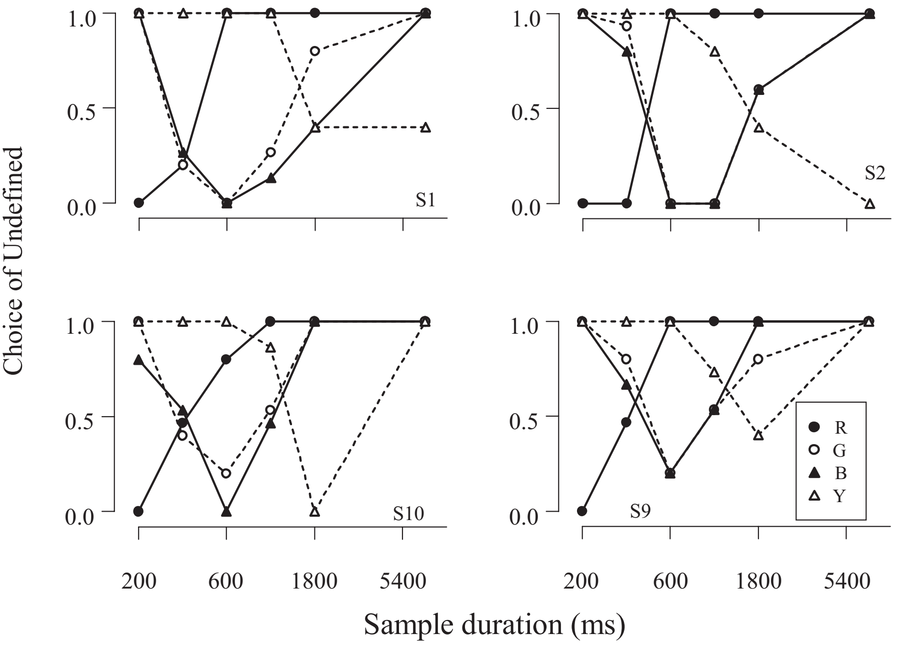

Fig. 2. The curves show the average proportion of choices of the undefined stimulus as a function of the sample duration, with the defined comparative stimulus (R, G, B, or Y) as a parameter. Sample duration is on a logarithmic scale.

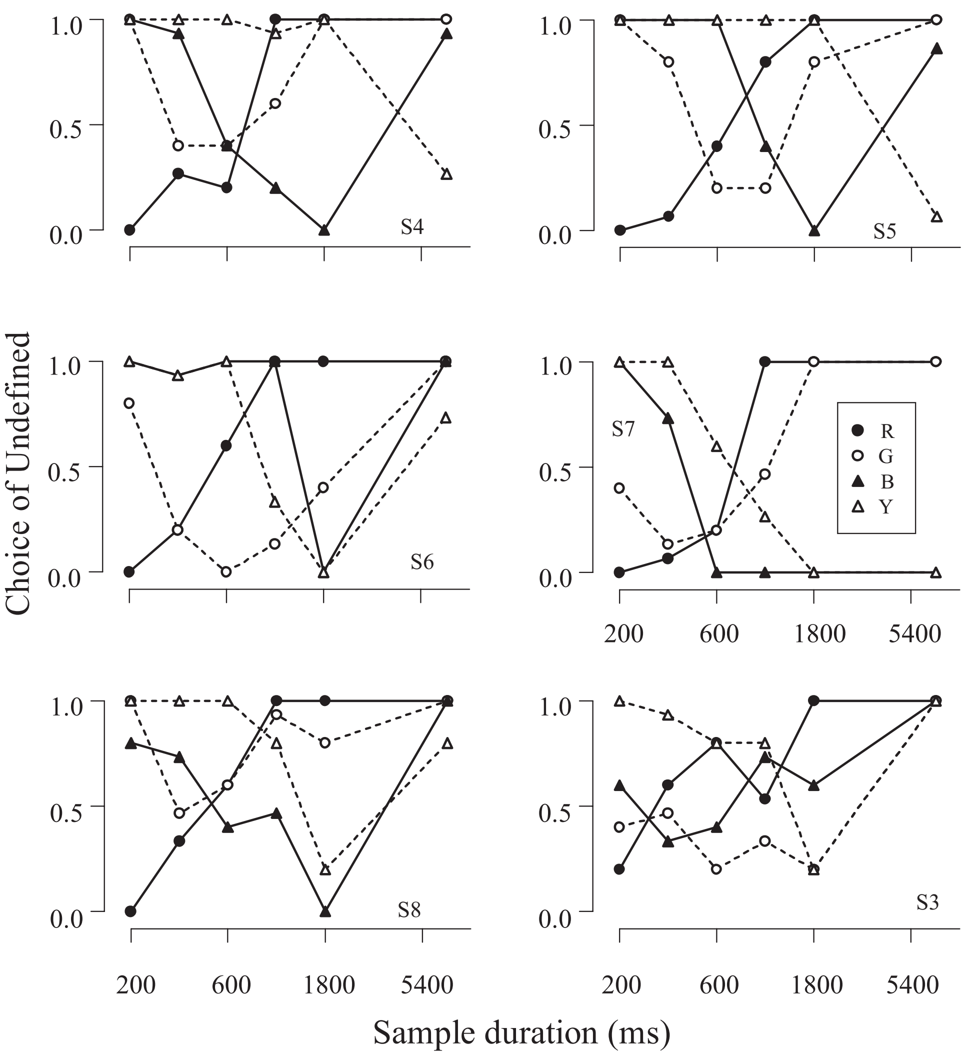

Fig. 3. The curves show the average proportion of choices of the undefined stimulus as a function of the sample duration, with the defined comparative stimulus (R, G, B, or Y) as a parameter. Sample duration is on a logarithmic scale.

for three participants (S6, S8, S3) it decreased and reached a minimum at 1800 ms, the defined sample, and then it increased again at the longest sample. Participant S7 performed similarly except that choice proportion did not increase at the longest sample. For the two remaining participants (S4 and S5), choice proportion decreased only at the longest sample of 7200 ms.

In summary, with some exceptions, the average results presented in Figure 1 represent well the individual results. All participants showed increasing curves in the tests with the R comparison. All participants showed U- or V-shaped curves in the tests with the G comparison. Eight of the 10 participants showed V- or U-shaped curves with the B comparison. These findings are consistent with our hypothesis. The data with the Y comparison were slightly more variable. Half of the participants showed V-shaped curves with higher values at the short samples, minimal values at 1800 ms, and higher values at the longest duration of 7200 ms, a choice pattern clearly consistent with the hypothesis. However, the other five participants showed curves with a minimum at 7200-ms samples, or with a minimum at 1800-ms samples but with an equally low value at the

7200-ms samples, a result inconsistent with our hypothesis.

Verbal Responses

The aforementioned results raise two questions. First, why was the B curve shifted to the right compared with the G curve for half of the participants? Second, why did the Y curve show higher variability at the longest samples compared with the other curves, especially the R and G curves? Potential answers to these questions can be gleaned from the interview conducted at the end of the exclusion tests. In the Appendix, we include the complete answers of all participants to the three questions asked by the experimenter. Here, we consider only the answers that refer to the choice of the comparison stimuli.

All participants explicitly stated a rule for responding by exclusion that was consistent with our initial hypothesis (i.e., the choice of the undefined stimulus was more likely when the duration of the sample moved away from the duration associated with the colored stimulus). In what follows we illustrate this finding for two participants:

[S10]: "If the beep was similar to the beep presented with the color, I clicked on the color. When I did not know how much time the beep had, because sometimes I had the sensation that I had never heard that one before, I clicked on the geometric figure."

[S8]: "I saw if the sound was equal to or different from the color. If it was equal, I chose the color. If the size of the sound was not associated with that color, I chose the figure."

Although the B and G comparisons were both associated to 600-ms tones, some participants stated that the B comparison was associated with a longer duration:

[S4]: "The times were increasing. First was the red, then the green was longer, then the blue was a little longer, and then the yellow."

[S5]: "It was more by the time. If I saw that the sound was short, I chose the red. If it was long, I went for the yellow. The green and blue confused me a bit. Sometimes I found that the green was shorter, sometimes not. But in general the green was smaller."

The previous statements suggest that some of the participants divided the comparisons into two classes according to the (subjective) duration of the corresponding sample. One class contained the R and G stimuli and was associated with shorter samples, and another class contained the B and Y stimuli and was associated with longer samples. Curiously, several participants seemed to have perceived the sample that corresponded to B as a bit longer than the sample that corresponded to G, a result that may help to explain the shift of the B curve to the right of the G curve.

However, there were several discrepancies between the verbal statements and the choice test results, and they suggest that the Y comparison could have been reallocated to the longest sample during testing. For example, participant S2 stated in the interview that "if the size of the sound was closer, more like the color, I clicked on the color. If it was very different from the color, I clicked on the figure." If we take into account that this participant discriminated the base durations reasonably well during training (overall percent correct = 77%), we would predict that in the exclusion tests the participant would prefer the Y stimulus after the 1800-ms training sample, but would prefer the undefined stimulus after the substantially longer 7200-ms untrained sample. However, this result did not occur (see Fig. 2). Thus, the verbal report of this participant is consistent with the R, G, and B curves but not with the Y curve, unless the Y stimulus is reallocated to the longest sample.

The cases of participants S4 and S5 are similar. Although they both performed well during training, with more than 79% correct choices after the 1800 ms sample, and verbally reported that they chose the undefined stimulus when "the sound associated with the color did not show up" (S4) or when the "[duration] did not resemble the one that I had associated" (S5), both always chose the undefined stimulus at 1800 ms and the Y stimulus at 7200 ms. In summary, for some participants, the Y stimulus associated during training with the longest (1800-ms) sample may have been reallocated to the longest (7200-ms) sample during testing.

Discussion

The goal of the present study was to investigate responding by exclusion using simple sample stimuli, that is, stimuli that varied along a single dimension. To that end, we studied

performance in conditional discriminations, matching-to-sample tasks that included samples (tones) differing only in duration. After the participants learned to associate specific sample durations with specific comparison stimuli (colored circles), they were tested with novel sample durations followed by novel comparison sets. The latter included always one of the comparisons defined during training (the R, G, B, or Y circles) and a new, undefined comparison (the U_1 to U_6 geometric shapes).

The average results showed that the probability of choosing the undefined stimulus when presented with the R comparison, previously associated with the 200-ms sample, was minimal at 200 ms and increased to 1 with the sample duration. When the undefined stimulus was presented with the Y comparison, previously associated with the 1800-ms sample, the opposite trend was observed, that is, preference for the undefined stimulus was strongest and close to 1 at the shortest durations and decreased, on average, at the longest 1800- or 7200-ms durations. When the undefined stimulus was presented with the G or B comparisons, preference for the undefined stimulus followed a U- or V-shaped curve (i.e., the probability of choosing the undefined stimulus was higher and close to 1 at the shortest and longest durations and lower at durations close to 600 ms, the sample previously associated with these comparisons).

These results reveal the accuracy of the trained discriminations because the R comparison was conditionally trained with the shortest sample, and the Y comparison was conditionally trained with the longest sample. With respect to the G and B comparisons, the similarity of the curve profiles indicate that the participants tended not to chose the undefined stimulus after the samples paired with G and B during training, and to choose it when the test sample were judged as sufficiently different from the samples paired with G and B during training. These results replicate the findings obtained with more complex stimuli that vary along multiple dimensions (e.g., Bates et al., 1979; Carey & Bartlett, 1978; Costa et al., 2001; Dixon, 1977; McIlvane et al., 1992; Stromer & Osborne, 1982; Wilkinson et al., 2009).

The individual data (Figs. 2 and 3) revealed differences among the participants. These differences are potentially interesting because they seem to be due to the interaction between responding by exclusion and temporal learning. The differences were most noticeable in the G, B, and Y functions (the R function remained roughly constant across participants, perhaps because no duration shorter than 200 ms was used). Concerning the G and B functions, most participants showed a U- or V-shaped function, but the location of the minimum was not always at 600 ms or adjacent values. For some participants the B function was to the right of the G function, with a minimum at 1800 ms. The relative positions of the B and G curves may express on the test trials the impression reported by some participants that the sample associated with B was longer than the sample associated with G. This misperception of the relative durations of the samples associated with G and B may stem from relational learning—because the 600-ms sample associated with G occurred in the context of the 200-ms sample associated with R, whereas the 600-ms sample associated with B occurred in the context of the 1800-ms sample associated with Y, the G sample may have been categorized as shorter than the B sample (see Machado et al., 2009 for the seemingly opposite context effect in pigeons, and Molet & Zental, 2008; and Spínola, Machado, de Carvalho, & Tonneau, 2013 for relational temporal learning).

As for the differences among participants in the Y function, they were restricted to the longest samples (the function remained roughly invariant for the shortest samples). For some participants the function was V- or U- shaped, but for others it remained high and then decreased at the longest sample. The difference may express a range restriction effect—had we used even longer test samples and perhaps all functions would have shown a V- or U-shaped form. In other words, perhaps the difference among participants is not the shape of the function, but the location of the minimum of the function. In this case the difference would be similar to the difference between the G and B functions. Alternatively, the difference among participants may express a framing effect during the test trials; because Y was associated to the longest sample during training, some participants may have reallocated Y to the longest sample during testing. Future studies must investigate these range effects in temporal learning and how they interact with responding by exclusion.

The procedure used in the present study differed from the procedures used in previous

studies also in the composition of the test trials. Typically, responding by exclusion is evaluated on trials in which the undefined comparison C_Y is presented together with a defined comparison, C_1, say, with C_1 being the S^+ for the sample S_1. With this arrangement of comparisons (C_1, C_Y) and an undefined sample, S_X, the participant excludes C_1 due to the previously defined relations (S_1:C_1) and thereby chooses C_Y the undefined stimulus. In contrast, in the present study, if for each sample (e.g., 350 ms) we order the four comparisons, starting with the comparison associated with the fewest choices of the undefined stimulus and ending with the comparison associated with the most choices of the undefined stimulus (e.g., in Fig. 1, for the 350-ms duration, the order would be R, G, B, and Y), then the order is consistent with the order predicted by the temporal-generalizations gradients. In other words, the average ordering is consistent with the idea that the choice of the undefined comparison C_Y varied inversely with the difference $|S_x–S_i|$, the difference between the duration of the test sample, S_x, and the duration of the sample associated with the defined comparison, S_i.

If the order is consistent with temporal generalization gradients, the absolute values of the choice proportions are not derivable exclusively from these gradients. Thus, having learned to choose R after 200-ms tones and to choose G after 600-ms tones, temporal generalization, as we commonly understand the concept, would not predict that preference for a new, undefined comparison over the G comparison would increase to 1 as the sample duration increased above 600 ms. Outside the basin of the generalization gradient, preference for the undefined stimulus over G would at most equal .5 (choice biases aside). We have argued that the modulation of the preference function by the sample durations as evinced in Figures 1 to 3 is caused by two processes, temporal generalization and responding by exclusion.

The foregoing conclusions based on the choice functions were generally consistent with the participants' verbal reports. Most participants (S2, S3, S4, S5, S8, and S9) said that when faced with a sample never seen before (temporal generalization), their choice was for the "different one", that is, the undefined stimulus (responding by exclusion). Some participants (S1, S6, and S7) chose the undefined stimulus when the defined comparison in the set was considered "shorter" (R, G) and the sample was considered longer. Similarly, they clicked the undefined stimulus when the sample was considered "short" and the defined comparison was B or Y. Notice that even in this case, the participants still chose the undefined comparison by excluding the defined comparison (one process, responding by exclusion) because the latter was not associated during training with the test sample (the other process, stimulus generalization).

The verbal reports were not always consistent with the participant's choices—a result also found when the correspondence between verbal and nonverbal behavior is directly investigated (e.g., Critchfield & Perone, 1990, 1993; Domeniconi, de Rose, & Perez, in press; St. Peter, Montgomery-Downs, & Massullo, 2012). However, analysis of verbal reports helped to clarify some features of the data, namely the fact that for half of the participants the B curve was to the right of the G curve, perhaps an expression of relational learning, and the fact that the Y curve showed significant variability at the longest sample, perhaps a framing effect. More generally, the analysis of verbal responses suggested specific factors that may influence responding by exclusion and that need further investigation.

The present study contributed to studies of exclusion by showing that responding by exclusion can take place with samples defined by duration, a stimulus modality not used in any other previous study. As for other modalities, the results showed that the undefined comparisons tended to be preferred to the defined comparisons after the new, undefined samples, but not after the trained, defined sample. The generality of responding by exclusion is thereby increased.

In addition, and more generally, the study also contributed by showing that responding by exclusion also occurs under the relatively simple condition in which the samples vary along a single dimension or property. With one-dimensional samples and equally simple comparisons, it is hard to conceive of an even simpler situation in which the phenomenon might take place. Hence, the present study may have defined some of the boundary conditions of responding by exclusion.

Responding by exclusion, that is, choosing an undefined comparison following an undefined sample, can take place under two distinct

conditions. The participant may reject the defined comparison or the participant may select the undefined comparison. These two conditions are accompanied by the same observable behavior but they imply two stimulus classes, one comprising the defined comparisons and the other comprising the undefined comparisons (McIlvane, Wilkinson, & de Souza, 2000; Wilkinson & McIlvane, 1997), and the relation between the sample stimuli and these two classes. In this light, we conclude by advancing a broader hypothesis concerning responding by exclusion. The hypothesis may be formulated in two logically equivalent ways, one positive and the other negative. The positive way states that to the extent that the undefined stimulus differs from the defined stimuli without introducing stimulus elements previously related to defined stimulus classes, responding by exclusion is likely to occur. These conditions may have been met in the present study because the undefined and defined stimuli differed only in duration. The negative way states that to the extent that the difference between undefined and defined stimuli introduces stimulus elements previously related to defined stimulus classes, responding by exclusion is less likely to occur. These conditions may have been met in the Costa, de Rose, and de Souza (2010) study, because adding suffixes to the undefined verbal samples (e.g., a diminutive, as in "mopadinho" instead of "mopadi") reduced responding by exclusion. Presumably because that lexical cue had been related during the participant's history of verbal behavior to objects of small size, that cue may have competed with the usual stimulus control sources of responding by exclusion. The same pattern of responses was present in Horst et al.'s (2011) study because previous exposure to the undefined object decreased the novelty of that object and, consequently, of responding by exclusion.

The present study investigated the minimal conditions required to obtain responding by exclusion. The use of stimulus samples that varied along a single dimension helped to quantify how "undefined" a new stimulus is. Its results expand our knowledge about the sources of stimulus control of responding by exclusion and may have implications for the development of educational procedures to teach language, for example, or for the design of methods to study responding by exclusion in nonverbal animals.

References

Aust, U., Range, F., Steurer, M., & Huber, L. (2008). Inferential reasoning by exclusion in pigeons, dogs, and humans. *Animal Cognition, 11*, 587–597.

Bates, E., Benigni, L., Bretherton, I., Camaioni, I., & Volterra, V. (1979). *The emergence of symbols: cognition and communication in infancy.* New York: Academic Press.

Beran, M. J. (2010). Use of exclusion by a chimpanzee (*Pan troglodytes*) during speech perception and auditory-visual matching-to-sample. *Behavioural Processes, 83*, 287–291.

Beran, M. J., & Washburn, D. A. (2002). Chimpanzee responding during matching to sample: control by exclusion. *Journal of the Experimental Analysis of Behavior, 78*, 497–508.

Call, J. (2006). Inferences by exclusion in the great apes: The effect of age and species. *Animal Cognition, 9*, 393–403.

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development, 15*, 17–29.

Costa, A. R. A., de Rose, J. C., & de Souza, D. G. (2010). Interferência de variáveis de contexto em sondas de exclusão com substantivos e verbos novos. *Acta Comportamentalia, 18*, 35–54.

Costa, A. R. A., Wilkinson, K. M., McIlvane, W. J., & de Souza, D. G. (2001). Emergent word-object mapping by children: further studies using the blank comparison technique. *Psychological Record, 51*, 343–355.

Critchfield, T. S., & Perone, M. (1990). Verbal self-reports of delayed matching to sample by humans. *Journal of the Experimental Analysis of Behavior, 53*, 321–344.

Critchfield, T. S., & Perone, M. (1993). Verbal self-reports about delayed matching to sample: Effects of the number of elements in a compound sample stimulus. *Journal of the Experimental Analysis of Behavior, 59*, 193–214.

Cumming, W. W., & Berryman, R. (1961). Some data on matching behavior in the pigeon. *Journal of the Experimental Analysis of Behavior, 4*, 281–284.

Dixon, L. S. (1977). The nature of control by spoken words over visual stimulus selection. *Journal of the Experimental Analysis of Behavior, 27*, 433–442.

Dixon, M. H., Dixon, L. S., & Spradlin, J. E. (1983). Analysis of individual differences of stimulus control among developmentally disabled children. In K. D. Gadow & I. Bialer (Eds.), *Advances in learning and behavioral disabilities* (vol. 2, pp. 85–110). Greenwich, CT: JAI Press.

Domeniconi, C., Costa, A. R. A., de Souza, D. G., & de Rose, J. C. (2007). Responder por exclusão em crianças de 2 a 3 anos em uma situação de brincadeira. *Psicologia Reflexão e Crítica, 20*, 342–350.

Domeniconi, C., de Rose, J. C. C., & Perez, W. F. (in press). Effects of correspondence training on self-reports of errors during a reading task. *The Psychological Record.*

Erdöhegyl, A., Topál, J., Virányl, Z., & Miklósi, A. (2007). Dog-logic: Inferential reasoning in a two-way choice task and its restricted use. *Animal Behaviour, 74*, 725–737.

Ferrari, C., de Rose, J. C. C., & McIlvane, W. J. (1993). Exclusion vs. selection training of auditory-visual conditional relations. *Journal of Experimental Child Psychology, 56*, 49–63.

Fields, L., Matneja, P., Varelas, A., Belanich, J., Fitzer, A., & Shamoun, K. (2002). The formation of linked perceptual classes. *Journal of the Experimental Analysis of Behavior, 78*, 271–290.

Floor, P., & Akhtar, N. (2006). Can 18-month-old infants learn words by listening in on conversations? *Infancy, 93*, 327–339.

Halberda, J. (2003). The development of a word-learning strategy. *Cognition, 87*, B23–B34.

Herman, L. M., Richards, D. G., & Wolz, J. P. (1984). Comprehension of sentences by bottle nosed dolphins. *Cognition, 16*, 129–219.

Horst, J. S., Samuelson, L. K., Kucker, S. C., & McMurray, B. (2011). What's new? Children prefer novelty in referent selection. *Cognition, 118*, 234–244.

Huziwara, E. M., Velasco, S. M., Tomanari, G. Y., de Souza, D. G., & Machado, A. D. (2012). Emergent relations in pigeons following training with temporal samples. *Learning & Behavior, 40*, 192–204.

Kaminski, J., Call, J., & Fischer, J. (2004). Word learning in a domestic dog: evidence for "fast mapping." *Science, 304*, 1682–1683.

Kastak, C. R., & Schusterman, R. J. (2002). Sea lions and equivalence: expanding classes by exclusion. *Journal of the Experimental Analysis of Behavior, 78*, 449–465.

Machado, A., Malheiro, M. T., & Erlhagen, W. (2009). Learning to time: a perspective. *Journal of the Experimental Analysis of Behavior, 92*, 423–458.

Machado, A., & Pata, P. (2005). Testing the scalar expectancy theory (SET) and the learning-to-time sample (LeT) in a double bisection task. *Learning & Behaviour, 33*, 111–122.

Marcicano, D. C., Carmo, J. S., & Prado, P. S. T. (2011). Software ProgMTS: Possibilidades de delineamento e condução de programas de ensino em Análise Experimental do Comportamento [programa de computador]. In proceedings of the 41st Annual Meeting of the Brazilian Society of Psychology: Training and Production of Knowledge in Psychology, Bethlehem.

McIlvane, W. J., Kledaras, J. B., Lowry, M. J., & Stoddard, L. T. (1992). Studies of exclusion in individuals with severe mental retardation. *Research in Developmental Disabilities, 13*, 509–532.

McIlvane, W. J., Kledaras, J. B., Munson, L. C., King, K. A., de Rose, J. C., & Stoddard, L. T. (1987). Controlling relations in conditional discrimination and matching by exclusion. *Journal of the Experimental Analysis of Behavior, 48*, 187–208.

McIlvane, W. J., Munson, L. C., & Stoddard, L. T. (1988). Some observations on control by spoken words in children's conditional discrimination and matching by exclusion. *Journal of Experimental Child Psychology, 45*, 472–495.

McIlvane, W. J., & Stoddard, L. T. (1981). Acquisition of matching-to-sample performances in severe mental retardation: learning by exclusion. *Journal of Mental Deficiency Research, 25*, 33–48.

McIlvane, W. J., Wilkinson, K. M., & de Souza, D. G. (2000). As origens da exclusão. *Temas em Psicologia, 8*, 195–203.

Molet, M., & Zentall, T. (2008). Relative judgments affect assessments of stimulus duration. *Psychonomic Bulletin & Review, 15*, 431–436.

Oliveira, L., & Machado, A. (2007). The effect of sample duration and cue on a double temporal discrimination. *Learning and Motivation, 39*, 71–94.

Pilley, J. W., & Reid, A. K. (2011). Border collie comprehends object names as verbal referents. *Behavioural Processes, 86*, 184–195.

Pruden, S. M., Hirsh-Pasek, K., Golinkoff, R. M., & Hennon, E. A. (2006). The birth of words: ten-month-olds learn words through perceptual salience. *Child Development, 77*, 266–280.

Saunders, K. J., & Spradlin, J. E. (1990). Conditional discrimination in mentally retarded adults: The development of generalized skills. *Journal of the Experimental Analysis of Behavior, 54*, 239–250.

Schloegl, C., Bugnyar, T., & Aust, U. (2009). Exclusion performances in non-human animals: from pigeons to chimpanzees and back again. In A. Blaisdell, L. Huber, S. Watanabe, A. Young, & Y. Yamazaki (Eds.), *Rational animals, irrational humans* (pp. 217–234). Tokyo: Keio University Press.

Schusterman, R. J., & Krieger, K. (1984). California sea lions are capable of semantic comprehension. *Psychological Record, 34*, 3–23.

Spínola, I., Machado, A., de Carvalho, M. P., & Tonneau, F. (2013). What do humans learn in a double, temporal bisection task: Absolute or relative stimulus durations? *Behavioural Processes, 95*, 40–49.

St. Peter, C. C., Montgomery-Downs, H. E., & Massullo, J. P. (2012). Improving accuracy of sleep self-reports through correspondence training. *The Psychological Record, 62*, 623–630.

Stromer, R. (1986). Control by exclusion in arbitrary matching-to-sample. *Analysis and Intervention in Developmental Disabilities, 6*, 59–72.

Stromer, R. (1989). Symmetry of control by exclusion in humans' arbitrary matching-to-sample. *Psychological Reports, 64*, 915–922.

Stromer, R., & Osborne, J. G. (1982). Control of adolescents' arbitrary matching-to-sample by positive and negative stimulus relations. *Journal of the Experimental Analysis of Behavior, 37*, 329–348.

Vieira de Castro, A. C., & Machado, A. (2010). Prospective timing in pigeons: Isolating temporal perception in the time-left procedure. *Behavioural Processes, 84*, 490–499.

Vieira de Castro, A. C., & Machado, A. (2012). The interaction of temporal generalization gradients predicts the context effect. *Journal of the Experimental Analysis of Behavior, 97*, 263–279.

Wearden, J. H. (1991). Human performance on an analogue of an interval bisection task. *The Quarterly Journal of Experimental Psychology, 43B*, 59–81.

Wearden, J., & Lejeune, H. (2008). Scalar properties in human timing: Conformity and violations. *The Quarterly Journal of Experimental Psychology, 61*, 569–587.

Wilkinson, K. M., de Souza, D., & McIlvane, W. J. (2000). As origens da exclusão. *Temas em Psicologia da SBP, 8*, 195–203.

Wilkinson, K. M., & McIlvane, W. J., (1997). Blank comparison analysis of emergent symbolic mapping by young children. *Journal of Experimental Child Psychology, 67*, 115–130.

Wilkinson, K. M., Rosenquist, C., & McIlvane, W. J. (2009). Exclusion learning and emergent symbolic category formation in individuals with severe language impairments and intellectual disabilities. *Psychological Record, 59*, 187–206.

Received: August 1, 2013
Final Acceptance: November 19, 2013

Appendix

At the end of the exclusion tests, the experimenter asked the participants the following three questions:

1) "Was there any criterion to click on one or the other circle? If yes, what was it?"
2) "What was your choice criterion when there was a colored circle and a figure that you had not seen before?" and
3) "Did you count to measure time? If so, how did you do it?"

The answers of each participant are presented below.

Participant S1

1. "At the beginning, I started clicking on the red, and then I continued clicking on the same color because the tone that was played was equal. Then came the green and the tone was played longer. When I did this, when the tone played one way, I clicked on one, and if it played another way, I clicked on the other."
2. "Then, when this started I did not understand. I first clicked on the figure that I never saw before, and nothing happened. Then I kind of determined that when the red or green appeared, those in the beginning were shorter, and a long sound was played, I clicked on the figure because red and green were short. When the sound was a little short and had blue and yellow, I also chose the figure."
3. "No, I did not count the time, I saw that it was short or long and clicked."

Participant S2

1. "I did this, not knowing if it was right, but subdivided the sound and compared it with one color or another in relation to the size of the sound. But each color did not always have the specific size. Like, I assigned that red was 0.5 to 1.0. The blue was 1.5 to 2.0. They varied in this interval that I determined, got it?"
2. "I saw it like this: If the size of the sound was close, it was more similar to a color, and I clicked on the color. If it was very different, I clicked on the figure."
3. "I counted. The sound itself had oscillations. I counted up to 14 oscillations as the longest, then I was seeing how many and compared them."

Participant S3

1. "There were sounds that were very small and others very big. Then there were the ones in the middle. The ones in the middle, I always got confused, but I paid attention to how long the sound was played and chose."
2. "Sometimes the sound was very similar to the one played before, and I clicked on the color. Then sometimes the sound was very different from the one that I heard before. When it was different from the one that I had listened to, I clicked on the figure, but I did not know if it was correct. I just thought that it was not the color so I clicked on the other one."
3. "Ah, it was fast, no time to count, no. Was I supposed to count?"

Participant S4

1. "The times were increasing. First was the red, then the green was longer, then the blue was a little longer, and then the yellow. In fact, I saw how long the sound played."
2. "When the sound that I had associated with the color did not appear, I clicked on the other one, the one that did not have color and had a figure inside."
3. "I was thinking 1, 2, 3, and counted in my head. But there were some that I could not, no way. When I tried, it was already done. I only counted the longer ones."

Participant S5

1. "I was going more by the time. If I saw that the sound was short, I chose the red. If it was long, I went for the yellow. The green and blue got me a little bit confused. Sometimes I thought that the green was shorter, sometimes not. But in general the green was smaller."
2. "I saw if the time that appeared corresponded to the time I had associated before with the color. If it was not similar to the one I had associated, I clicked on that figure that had never appeared before."
3. "Ah, yes, I counted. In my head, when the tone started I started 1, 2, 3, and so on."

Participant S6

1. "It was more the duration of the sound that mattered to me, and I made the choice. It was like the sound was growing. Started with red, growing to green, then blue, and then yellow."
2. "If the time was the one that I remembered, I clicked on the color. There was the short group that was red and green and the longer one that was blue and yellow. If the sound was big, for example, and had green and the figure, I clicked on the figure, got it? The most difficult part was when it had yellow and the sound was big."
3. "I counted nothing. I only felt if the sound was fast or slow."

Participant S7

1. "I thought that the red was a little bit fast, the green a little bit longer than the red, and then came the blue and then the yellow. Then later I compared the two colors that appeared. If it was very short, it was red. If it was short, it was green. If it was a little longer, it was blue. If it was very long, it was the yellow."
2. "At the beginning, I started realizing that when the sound was very long, the yellow and a figure like a star appeared. Then I always clicked on the star because the yellow was not that long. Then after this no longer worked, I realized that there were shorter and longer sounds. If the short sound played and green or red appeared, I pushed the color because red and green were a small pair. When green or red appeared and the sound was longer, I clicked the figure because I thought that it [the tone] was not of the small pair green/red but rather of the blue/yellow."
3. "I saw that sometimes when the sound played, it had some little parts higher, like little jumps. Then I counted that. Sometimes I could not do that, but sometimes I could."

Participant S8

1. "By my logic, it was the duration of the sound that determined if I chose one or the other. If the sound was long, I pushed blue or yellow, if it was faster, then it was green or red. And between the red and green, the red was fastest, and between the blue and yellow, the yellow was slower."
2. "I saw if the sound was equal to or different from the color. If it was equal, I chose the color. If the size of the sound was not associated with that color, I chose the figure."
3. "I did not count because it was too fast; one couldn't count."

Participant S9

1. "The red was fast, the green a little less fast, the blue even less fast, and the yellow was longer. In fact, the red was linked to the green and was shorter, and the blue was linked with the yellow and was longer."
2. "If the time was different from the one that I remembered at the beginning when I clicked on the color, then I chose the little figure. If the time corresponded to a color, I clicked on the color."
3. "At the beginning, I counted in my head, but after too many sounds began to appear, I couldn't count anymore."

Participant S10

1. "I 'gave' a time to each color. For example, the red had 1 second, the yellow had between 3 and 4 seconds, and the green and blue were in between. I could not distinguish between the green and blue, but if it was a little bit longer, it was blue because it was together with the yellow at the beginning."
2. "If the beep was similar to the beep presented with the color, I clicked on the color. When I did not know how much time the beep had, because sometimes I had the sensation that I had never heard that one before, I clicked on the geometric figure."
3. "I tried to count in my head how many seconds elapsed. It was not exactly the correct seconds, but I could count."