Behavioural Processes 95 (2013) 40–49

Contents lists available at SciVerse ScienceDirect

Behavioural Processes

journal homepage: www.elsevier.com/locate/behavproc

What do humans learn in a double, temporal bisection task: Absolute or relative stimulus durations?

Igor Spínola, Armando Machado, Marilia Pinheiro de Carvalho*, François Tonneau

University of Minho, Portugal

ARTICLE INFO

Article history:
Received 29 September 2012
Received in revised form
24 December 2012
Accepted 13 January 2013

Keywords:
Relative versus absolute
Double bisection task
Psychometric function
Timing
Humans

ABSTRACT

The relative-coding hypothesis of temporal discrimination asserts that humans learn to respond to the relative duration of stimuli ("short" and "long"). The most frequently used procedure to test the hypothesis is the double bisection task. In one task, participants learn that red and green are the correct comparisons following 2 s (short) and 5 s (long) samples respectively. In another task, participants learn that triangle and circle are the correct comparisons following 3.5 s (short) and 6.5 s (long) samples, respectively. Later the samples of one task are tested with the comparisons of the other task, and vice versa. According to the hypothesis, participants will choose red following a 3.5 s sample because that sample is short and red is the comparison that goes with short. Similarly, they will choose circle following 5 s samples because that sample is long and circle goes with long. We replicated this procedure and improved it by introducing several sample durations during testing to obtain the whole psychometric function of each task. Results from Experiment 1 only partially corroborated the relative-coding hypothesis. Results from Experiment 2 did not corroborate the hypothesis. The combined data from Experiments 1 and 2 partially corroborate the hypothesis. Alternatively, we present an explanation of relative-coding-like results that posits exclusively absolute coding of temporal stimuli.

This article is part of a Special Issue entitled: SQAB 2012.

1. Introduction

The nature of the discriminations learned in temporal bisection tasks has been a matter of debate: Do subjects learn to discriminate the absolute or the relative duration of the stimuli? Different procedures and different measures have been employed to clarify this issue (e.g., Carvalho and Machado, 2012; Church and Deluty, 1977; Hulse and Kline, 1993; Maia and Machado, 2009; Molet and Zentall, 2008; Zentall et al., 2004). One such procedure was first proposed by Zentall et al. (2004) for the study of timing with pigeons. Later, Molet and Zentall (2008) adapted this task to human participants.

Molet and Zentall (2008, Experiment 2) used the following double bisection task: in one bisection task, participants learned to choose a red comparison stimulus following a 2 s sample, and to choose a green comparison stimulus following a 5-s sample. In this discrimination, 2 s is the relatively "short" sample whereas 5 s is the relatively "long" sample. In another bisection task, the participants learned to choose a triangle following a 3.5 s sample, and to choose a circle following a 6.5 s sample. Here, 3.5 s is relatively "short" and 6.5 s is relatively "long". The authors evaluated a *relative-coding hypothesis* according to which during the training of each bisection task the participants learn that 2 s and 3.5 s are both "short" and that 5 s and 6.5 s are both "long" stimuli. Also, the hypothesis holds that participants learn to respond to the relative value of the samples according to the following diagram: short → red, long → green, short → triangle, long → circle.

The rationale for Molet and Zentall's (2008) procedure relies on a common finding in human timing: the bisection point or the point of subjective equality tends to equal the arithmetic mean of the sample durations used during training (e.g., Wearden, 1991; Wearden and Ferrara, 1995; Wearden et al., 1997). That is, following a sample that equals the arithmetic mean, participants choose each of the available comparisons with the same probability. Now, notice that in Molet and Zentall's study (2008, Experiment 2) the 3.5 s stimulus, the short sample in the second task, is the arithmetic mean of the sample durations used in the first task. Hence, if regular timing results occurred, participants would be indifferent between the 'red' and 'green' comparisons following a 3.5 s test sample. Similarly, because 5 s, the long sample in the first task, is the arithmetic mean of the sample durations used in the second task, participants should be indifferent between the triangle and circle comparisons following a 5 s test sample. However, the relative-coding hypothesis predicts different results. If participants learned that 3.5 s is "short", they should choose red when given the opportunity to choose between red and green, because red is a comparison

* Corresponding author at: Escola de Psicologia, Universidade do Minho, 4710-057 Braga, Portugal. Tel.: +351 253 601 396.
E-mail address: marilia.pinheiro@yahoo.com.br (M.P. de Carvalho).

0376-6357/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.beproc.2013.01.003

associated with "short". Similarly, if participants leaned that 5 s is "long", they should choose the circle when given the opportunity to choose between a triangle and a circle because the circle is a comparison associated with "long". In sum, according to the relative-coding hypothesis, participants should group stimuli that share the same relative value (either "short" or "long") by responding similarly to them.

Since its publication, the original study by Zentall et al. (2004) has been replicated, producing evidence in favor of both relational discriminations (Molet and Zentall, 2008; Zentall et al., 2004) and absolute discriminations (Maia and Machado, 2009). Though the results reported by Zentall et al. (2004) and later by Molet and Zentall (2008) raise the possibility that animals and humans encode temporal intervals relationally, the original arguments used by the authors to explain the relative discrimination of temporal intervals were not sufficiently developed. In fact, Zentall et al. (2004) asserted that subjects learned "[. . .] to treat the 3.5 s sample and the 2 s sample similarly, [and] to treat the 5 s sample and the 6.5 s sample similarly" (pp. 113–114), but they did not specify what "treating stimuli similarly" meant. Subsequently, Maia and Machado (2009) elaborated Zentall et al.'s hypothesis: During training subjects learn to respond to the absolute value of stimuli, but they may also learn to encode the 3.5 s and the 2 s as 'short' and the 5 s and the 6.5 s as 'long'. Then, during testing, subjects choose the comparisons on the basis of its associations with the codes 'short' and 'long'. That is, the relative code would work as a mediator response between the stimulus duration and the choice of the comparison stimuli.

Furthermore, Zentall et al. (2004) and Molet and Zentall (2008) did not discuss a number of procedural issues that limit our understanding of relational timing in bisection tasks. One outstanding procedural issue is that both the studies by Molet and Zentall (2008) and by Zentall et al. (2004) tested relative-coding effects with two sample durations only: In the case of Zentall et al.'s study they used the geometric means of the trained durations; in the case of Molet and Zentall's study they used either the geometric means (Experiment 1) or the arithmetic means (Experiment 2). As a consequence, their conclusions on the relative coding of temporal intervals arose from an interpretation of only one point of a whole psychometric function. Furthermore, testing with only the arithmetic means does not provide some relevant measures of timing behavior (e.g., the slope of psychometric functions). A final weakness in the studies by Molet and Zentall (2008) and Zentall et al. (2004) is that they did not test whether the bisection points actually occurred at the arithmetic means. Inferring shifts in the bisection points from arithmetic means without knowing empirically where the bisection points were originally located is a risky decision.

Our study deals with these pending issues about relative timing. In Experiment 1, we tried to replicate the results found by Molet and Zentall (2008) with human participants, but we introduced several testing durations in order to evaluate a whole psychometric function. In Experiment 2, we conducted tests at two different moments, so that we could make more reliable conclusions about shifts in the bisection points and about shifts in the psychometric functions. Finally, we attempt to provide an alternative explanation of apparent relative coding by positing exclusively absolute coding of temporal stimuli.

2. General method

2.1. Apparatus

Experimental sessions were run in a small quiet room equipped with an ASUS L5800 Notebook PC. Black, yellow (both 161 × 102 pixels), red and green rectangles (both 135 × 51 pixels), as well as a dark-gray triangle and a dark-gray circle (both 90 × 90 pixels) served as stimuli and were presented against a light-gray colored computer screen with a pixel resolution of 1400 × 1050. Responses were given through a standard computer mouse. A Visual Basic 2005 (Microsoft®) program was used to control all experimental events and record data.

2.2. Procedure

Each participant was trained and tested in one session that lasted from 45 to 55 min. During the session, the participant was exposed to two different tasks, one involving 2 s and 5 s samples (Range 1), the other involving 3.5 s and 6.5 s samples (Range 2). For clarity purposes we will call the trained sample durations *standard durations*. The training order was counterbalanced across participants so that half of the participants first learned Range 1, whereas the other half first learned Range 2. After or during training (see below), the participants performed one or more stimulus generalization tests with temporal stimuli other than the standards.

Training and testing involved delayed-matching-to-sample trials. At the beginning of a trial, a yellow rectangle (6 cm wide × 4 cm high) that stood for the sample stimulus appeared at the center of the computer screen. The yellow rectangle lasted for the duration of a temporal interval. After the sample had elapsed, the rectangle turned black and a pair of comparison stimuli was presented at the bottom of the screen, one on the left side and one on the right side. The participant had to choose one of the comparison stimuli by clicking on it with the mouse cursor. Then the comparison stimuli were turned off. Each correct choice was rewarded during training trials by adding 1 point to the number of points earned by the subject, and the total number of points already earned was displayed in a window at the bottom of the screen. Incorrect choices (as well as choices during test trials) were never rewarded. In all cases, a 2 s ITI signaled by the black rectangle followed each trial.

As shown in Fig. 1, training trials for Range 1 used 2 s ("short 1", or S_1) and 5 s ("long 1", or L_1) samples, whereas training trials for Range 2 used 3.5 s (S_2) and 6.5 s (L_2) samples. Testing trials within Range 1 employed stimuli of 2.8 s, 3.5 s and 4.3 s, while testing trials of Range 2 employed stimuli of 4.3 s, 5 s and 5.8 s. Notice that the sample stimulus of 3.5 s employed in testing within Range 1 equals the arithmetic mean of the standard durations for this range; similarly, the sample stimulus of 5 s employed in testing within Range 2 equals the arithmetic mean of the standard durations for this range.

For half of the participants, the comparison stimuli presented during Range 1 training were a red and a green rectangle (5 cm wide × 2 cm high), and the pair of stimuli presented during Range 2 training were a black triangle and a black circle. The opposite assignment held for the other half of the participants. For clarity however, from now on we will describe the procedure as if all the participants had learned the assignments, '2 s → red, 5 s → green' and '3.5 s → triangle, 6.5 s → circle' (Fig. 1). The position (left or right) of the comparison stimuli on each trial was random. In all cases, the red rectangle and the black triangle were the correct comparisons following the "short" standards, whereas the green rectangle and the black circle were the correct comparisons following the "long" standards.

2.3. Experiment 1

This experiment replicated the procedure employed in Molet and Zentall's (2008) second experiment. Our research question was the same as theirs: While being trained on temporal bisection tasks, do humans learn the discrimination on the basis of the relative value of the stimuli ("short" and "long"), and afterwards tend to respond similarly to different stimuli that share the same relative value? In other words, when presented with standard S_2, will the participants choose the comparison previously associated

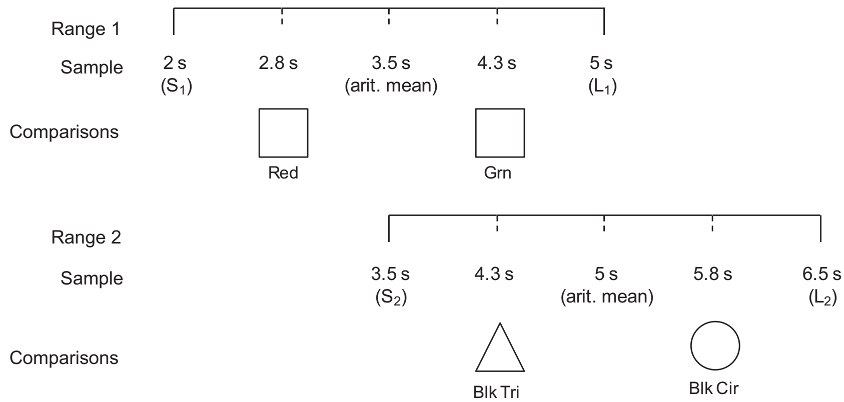

Fig. 1. Training (solid lines) and testing (dashed lines) stimuli used in the experiments. 'S_1' and 'L_1' stand respectively for the "short" and "long" standards in Range 1. 'S_2' and 'L_2' stand respectively for the "short" and "long" standards in Range 2. 'Red' and 'Grn' stand respectively for the red and green comparisons used with Range 1. 'Blk Tri' and 'Blk Cir' stand respectively for the black triangle and black circle comparisons used with Range 2. For illustrative purposes, the figure shows only one assignment of the comparison pairs to ranges 1 and 2.

Table 1
Experimental design of Experiment 1.

	Group 1	Group 2
Phase 1 – Training	2.0 s vs. 5.0 s (Range 1)	3.5 s vs. 6.5 s (Range 2)
Phase 2 – Training	3.5 s vs. 6.5 s (Range 2)	2.0 s vs. 5.0 s (Range 1)
Phase 3 – Training	Mixed Range 1 and Range 2	
Phase 4 – Testing	Mixed training and testing: Range 1 and Range 2	

with standard S_1, thereby showing that they have grouped the two "short" standards under the same response? Additionally, when presented with standard L_1, will the participants choose the comparison previously associated with standard L_2, thereby showing that they have grouped the "long" standards under the same response? The major difference between our experiment and Molet and Zentall's is that we tested several intermediate stimulus durations in the test phase instead of testing only the arithmetic mean of the standards.

2.3.1. Participants

Sixteen female and four male undergraduate students from the University of Minho were recruited by means of e-mail advertisement (median age: 22 years, range: 18–33). None of them had prior experience with bisection tasks or any other type of timing experiment. The participants were randomly and equally distributed between two groups that differed only in the order of training the two ranges.

2.3.2. Procedure

As shown in Table 1, the procedure was divided into four phases. Phases 1–3 involved training the temporal discriminations. Phase 4 involved testing.

Phase 1 – Training the first discrimination. After sitting in front of the computer the participant read the following message[1]: "*Thank you for participating in our study on timing. You will see a yellow light that will come on for some period of time. The yellow light will be followed by a choice between two buttons. For each of two durations of the yellow light you will have to figure out which button is the correct one. The position of the buttons is not important; it is the color or the geometric shape of the buttons that matters. If you choose correctly, you will be given a point on the screen. Please do not count or perform rhythmic activities during the experiment*". At the bottom of the screen there was a button that displayed the message "*Click here to start*". After clicking the button, the experiment proper began.

There was a minimum of 40 training trials comprising 20 trials with a "short" standard as sample (either 2 s or 3.5 s, depending on the range being trained) and 20 trials with a "long" standard (either 5 s or 6.5 s). The presentation order of the two types of trials was randomized. Training proceeded with correction: any particular trial was repeated until a correct choice was made. Only choices made on the first presentation of each trial were considered for data analysis. Phase 1 lasted until participants reached a criterion of 7 correct trials out of 9 trials for each of the two standards.

Phase 2 – Training the second discrimination. Right after the criterion for Phase 1 was met the following instructions appeared on the screen: "*Good! Now there will be a new problem with new durations. Once again, you will have to figure out which button is the correct one for each duration. Please do not count or perform rhythmic activities during the experiment*". After clicking on the "*Click here to start*" button, Phase 2 began. This phase was identical to Phase 1, except that the participants first trained with Range 1 were now trained with Range 2 and vice versa. There was a minimum of 40 trials, 20 for the "short" standard and 20 for the "long" standard. The learning criterion was the same as in Phase 1.

Phase 3 – Mixed training. Right after the criterion for Phase 2 was met the following instructions appeared on the screen: "*Now we are going to combine the two problems to see if you can perform both of them at the same time. Please do not count or perform rhythmic activities during the experiment*". After clicking on the "*Click here to start*" button, 40 trials were presented, 10 for each of the four standard durations previously trained. Phase 3 lasted until a criterion of 4 correct trials out of 5 was reached for each of the four standards.

Phase 4 – Stimulus generalization testing. After Phase 3, the following instructions appeared on screen: "*Now we are going to continue, but you will not receive feedback after your choices. You will still be given points, but they will not appear on the screen. Please do not count or perform rhythmic activities during the experiment*". After clicking on the "*Click here to start*" button, testing began. Phase 4 had 40 training trials similar to the trials of Phase 3 (10 trials for each of the four standards) and 60 testing trials, 10 for each of the testing durations. On 10 of the 60 testing trials, the 3.5 s standard was displayed along with the pair of comparison stimuli that had previously been used during training the 2 s vs. 5 s task. On another 10 of the 60 testing trials, the 5 s standard was displayed along with the pair of comparison stimuli that had been used during training

[1] All instructions have been translated from European Portuguese.

the 3.5 s vs. 6.5 s task. At the end of Phase 4, we asked participants in a brief interview whether they had counted or performed any rhythmic activity during the experiment.

2.3.3. *Results and discussion*

We analyzed the proportion of choices of the comparisons associated with the "Long" standards as a function of sample duration. We will call such choices "Long" responses. Thus, choosing 'green' and 'circle' were the "long" responses in Ranges 1 and 2 respectively.

2.3.3.1. Acquisition. Data from three participants revealed proportions of "Long" responses to the "short" standards that were greater than 0.5 or proportions of "Long" responses to the "long" standards that were lesser than 0.5. Though the acquisition criterion did not predict a maximum number of training trials, we considered that these three participants failed to acquire the baseline discriminations. Because of that, their data were excluded from further data analysis.

In Phase 1, 8 out of 10 participants trained with Range 1 and 9 out of 10 participants trained with Range 2 reached the learning criterion within the required minimum of 40 trials. In Phase 2, 7 out of 10 participants trained with Range 1 and 9 out of 10 participants trained with Range 2 reached the learning criterion within the required minimum of 40 trials.

Mann–Whitney tests revealed that the number of trials needed to learn Range 1 in Phase 1 (median = 40 trials) did not differ significantly (U = 46, *ns*) from those needed to learn it in Phase 2 (median = 40 trials). Similarly, the number of trials needed to learn Range 2 in Phase 1 (median = 40 trials) did not differ significantly (U = 49.5, *ns*) from those needed to learn it in Phase 2 (median = 40 trials). This suggests that there was no order effect on learning each range. A Wilcoxon signed rank test revealed no significant difference ($Z = -1.07$, *ns*) between learning Range 1 first (median = 40 trials) and learning Range 2 second (median = 40 trials) or between learning Range 2 first (median = 40 trials) and learning Range 1 second (median = 40 trials). This suggests that the number of trials required for acquiring the discriminations was independent of their ranges.

2.3.3.2. Testing. Fig. 2 shows the group averaged psychometric functions obtained for each range. A visual inspection of the functions reveals that the proportion of "Long" responses increased monotonically with sample duration. This pattern of behavior resembles those reported in previous bisection studies with humans (e.g., Allan and Gibbon, 1991; Wearden, 1991; Wearden and Ferrara, 1995, 1996).

One-way repeated measures ANOVAs revealed a significant effect of duration for both Range 1 [$F(4,64) = 87.34$, $p < .0001$] and Range 2 [$F(4,64) = 68.73$, $p < .0001$], confirming that participants emitted more "Long" responses as sample duration increased.

For comparison purposes, we performed Molet and Zentall's (2008) analyses on our data. First we calculated the mean proportion of "Long" responses during testing trials for both the 3.5 s and the 5 s samples, obtaining values of P ("Long") equal to 0.429 and 0.647 respectively (Table 2). One-sample t-tests showed that P ("Long") for the 3.5 s sample was not statistically different from

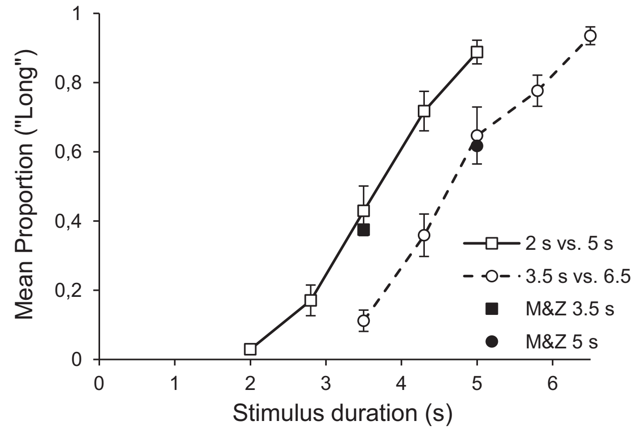

Fig. 2. Psychometric functions obtained in Experiment 1. The empty dots represent data from the present study; filled dots (at the arithmetic means) represent data reported by Molet and Zentall (2008). Standard error bars are displayed for each sample duration.

chance [$t(16) = -0.99$, $p = .17$, one-tailed], whereas P ("Long") for the 5 s sample significantly differed from chance [$t(16) = 1.79$, $p = .05$, one-tailed]. These results indicate that the participants did not group the "short" standards from both tasks under the same code; nevertheless, participants may have grouped the "long" standards from both tasks under the same code, as suggested by Molet and Zentall (2008). In terms of individual data, 12 out of the 17 participants presented a proportion of "Long" responses above 0.5 for the 5 s sample.

Molet and Zentall (2008) argued that averaging the data from 3.5 s trials and 5 s trials for each participant would correct for possible deviations from the arithmetic mean and that this averaging would provide a better test of the relative-coding hypothesis. Hence, we performed a one-sample t-test on the averaged data and found that the mean proportion of choices of the comparisons associated with S_1 and L_2 equalled 0.609, which was not significantly different from chance [$t(16) = 1.68$, $p = .06$, one-tailed]. Though not significant, the average response proportion was greater than 0.5 in 11 out of 17 participants.

In addition to the proportion of "Long" responses at arithmetic means, the psychometric functions for both temporal ranges enabled us to obtain additional information, including bisection points and slopes. Table 2 shows the average results.

Individual bisection points were determined by linear interpolation (Wearden and Ferrara, 1995, 1996) and then averaged. A one-sample t-test revealed that the mean bisection point of Range 1 ($M = 3.75$; SE = 0.15) was not significantly different from the arithmetic mean of 3.5 s [$t(16) = 1.64$, $p = .06$, one tailed]. Another one-sample t-test revealed that the mean bisection point of Range 2 ($M = 4.69$; SE = 0.16) was significantly below the arithmetic mean of 5 s [$t(16) = -1.92$, $p = .037$, one tailed]. This later result is consistent with the relative-coding hypothesis: If the participants learned that 5 s is a "long" stimulus, and thus, emitted more "Long" than "Short" responses in the presence of a 5 s sample, the corresponding psychometric function should cross the horizontal line of 50% to the left of 5 s and show a bisection point below the arithmetic mean (dotted

Table 2
Summary statistics for Experiment 1. P ("Long") at arithmetic mean is the mean proportion of "Long" responses when the sample duration equaled the arithmetic mean. L95% and U95% are respectively the lower and the upper limits of the 95% confidence interval for the mean.

Discrimination	P ("Long") at arithmetic mean			Bisection points (s)			Slope of the psychometric function
	Mean	L95%	U95%	Mean	L95%	U95%	
Range 1	0.43	0.30	0.55	3.75	3.19	4.35	0.15
Range 2	0.65	0.503	0.79	4.69	4.4	4.97	0.17

curve in Fig. 2). However, although consistent with the relative-coding hypothesis, this result might just as well be assigned to the long/short standards ratio employed. When the L/S ratio falls below 2:1, the bisection point tends to be closer to the geometric mean of the standards (Allan and Gibbon, 1991; Wearden and Ferrara, 1996). Since the long/short ratio in the discrimination between 3.5 s vs. 6.5 s is 1.86, the bisection point is expected to be closer to the geometric mean (Range 2 geometric mean = 4.77 s).

The slopes of the psychometric functions for both temporal ranges were estimated by dividing the difference limen of each psychometric function by its bisection point, yielding the Weber ratio for each curve. The obtained Weber ratios were 0.15 (median = 0.13) and 0.17 (median = 0.13) for Range 1 and Range 2 respectively. A Wilcoxon signed rank test revealed no significant differences between the Weber ratios ($Z = -0.22$, *ns*), suggesting that there were no differences in temporal sensitivity between the temporal ranges employed.

Overall the results of Experiment 1 provide mixed evidence regarding the relative-coding hypothesis put forward by Molet and Zentall (2008). On one hand, both the mean proportion of "Long" responses at the 5 s sample and the bisection point of the psychometric function for Range 2 provide statistically significant evidence consistent with the assumed grouping of the "long" standards. On the other hand, no statistically significant evidence was found to support a grouping effect with respect to the "short" standards, even though the trend of the results was in the same direction as that reported by Molet and Zentall (2008). (Interestingly, Zentall et al. (2004) also concluded in favor of the relative-coding hypothesis on the basis of partially favorable results. Their evidence for grouping occurred with respect to the "short" standards).

Molet and Zentall (2008) justified their relational-coding hypothesis by assuming both that the bisection point for human participants lies at the arithmetic mean of the standards, and that deviations from arithmetic-mean bisection involved a grouping of temporal stimuli based on their relative duration (either "short" or "long"). However, Molet and Zentall did not evaluate directly whether the bisection point was indeed at the arithmetic mean. As mentioned before, although arithmetic-mean bisection is the most common finding in human timing, it is not ubiquitous. The participants in Molet and Zentall's study might have shown deviations from arithmetic-mean bisection even before the second task. If this was the case, the relative-coding hypothesis would be seriously challenged. Reliable conclusions based on changes in the bisection point or the shape of the psychometric function for the first task can be reached only by measuring them before and after training on the second task. Experiment 2 was conducted to address this issue.

2.4. Experiment 2

This experiment was designed so that we could evaluate the effects of the second task on the psychometric function from the first task. To this end, we obtained a psychometric function pertaining to the first task both before and after training on the second task, and then compared the two functions.

2.4.1. Participants

Nineteen students and a staff member from the University of Minho (7 women and 13 men) participated. Ages ranged from 18 to 47 years (median: 20). No participant had prior experience with bisection tasks.

2.4.2. Apparatus

Same as in Experiment 1.

Table 3
Experimental design of Experiment 2.

	Group 1	Group 2
Phase 1 – Training	2 s vs. 5 s (Range 1)	3.5 s vs. 6.5 s (Range 2)
Phase 2 – Testing	Range 1	Range 2
Phase 3 – Training	3.5 s vs. 6.5 s (Range 2)	2 s vs. 5 s (Range 1)
Phase 4 – Training	Mixed Range 1 and Range 2	
Phase 5 – Testing	Mixed training and testing: Range 1 and Range 2	

2.4.3. Procedure

As in Experiment 1, we used separate training phases for each range (Phases 1 and 3 in Table 3), a mixed training phase with both ranges (Phase 4) and a final testing phase with both ranges (Phase 5). However, in contrast to Experiment 1 a testing phase (Phase 2 in Table 3) was added after learning the first range. Here we describe the testing phase that was introduced in Experiment 2. All other experimental details remained unchanged.

Phase 2 – Generalization test for the first task. After meeting the learning criterion in the first task (Phase 1), the following message appeared on the computer screen: "*Now we are going to continue, but you will not be given feedback. You will still be given points, but they will not appear on the screen. Please do not count or perform rhythmic activities during the experiment*". After reading the instructions, the subject clicked on the "*Click here to Start*" button and the test began. Each participant was tested with respect to the temporal range learned in Phase 1, that is, half of the participants were tested with Range 1 and the other half was tested with Range 2 (see Fig. 1). This phase had 40 trials: 8 training trials for each standard duration, and 8 testing trials for each of the remaining sample durations.

2.4.4. Results and discussion

2.4.4.1. Acquisition. During Phase 1, 8 out of 10 participants trained with Range 1 and 9 out of 10 participants trained with Range 2 met the learning criterion within the required minimum of 40 trials. In Phase 3, 7 out of 10 participants trained with Range 1 and 9 out of 10 participants trained with Range 2 reached the learning criterion within the required minimum of 40 trials. A Mann–Whitney test showed that the number of trials needed to learn Range 1 in Phase 1 (median = 40 trials) did not differ significantly from those needed to learn it in Phase 3 (median = 40), $U = 40$, *ns*. In the same way, the number of trials needed to learn Range 2 in Phase 1 (median = 40) did not differ significantly from those needed to learn it in Phase 3 (median = 40), $U = 45$, *ns*. This suggests that no order effect occurred with respect to learning the baseline temporal discriminations. A Wilcoxon signed rank test revealed no significant difference between learning Range 1 first (median = 40) and learning Range 2 s (median = 40) ($Z = -1$, *ns*), or between learning Range 2 first (median = 40) and learning Range 1 s (median = 40) ($Z = -1.34$, *ns*). This suggests that the acquisition of both tasks was independent of the absolute durations used in each range. Accordingly, acquisition results were similar to the ones obtained in the first experiment.

2.4.4.2. Testing in Phase 5. The same analysis from Experiment 1 was carried out on the test data obtained after both temporal tasks had been learned in Experiment 2. Fig. 3 shows the group averaged psychometric functions for each range. One-way repeated measures ANOVAs revealed a significant effect of duration on the Range 1 function [$F(4,64) = 116.52$, $p < .0001$] and on the Range 2 function [$F(4,64) = 76.59$, $p < .0001$], confirming that the proportion of "Long" responses increased with sample duration.

Regarding the proportion of "Long" responses at the arithmetic means, one-sample *t*-tests showed that the mean proportion of "Long" responses at the 3.5 s sample ($M = 0.43$) [$t(16) = -1.04$, $p = .16$, one-tailed] and the mean proportion of "Long" responses at the 5 s sample ($M = 0.56$) [$t(16) = 0.82$, $p = .21$, one-tailed] were not

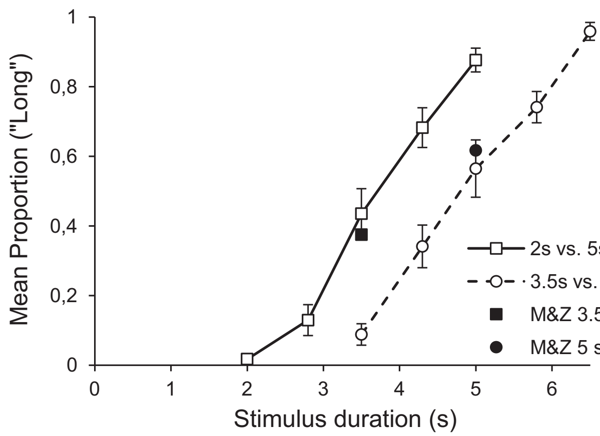

Fig. 3. Psychometric functions obtained in the final testing phase of Experiment 2. Standard errors are displayed for each sample duration.

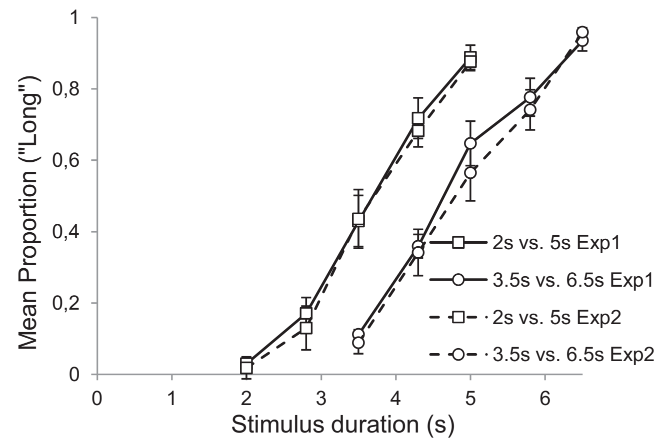

Fig. 4. Psychometric functions from Experiment 1, Phase 4, and Experiment 2, Phase 5. Standard errors are displayed for each sample duration.

significantly different from chance. This suggests that a grouping effect based on relative values occurred neither for the "short" nor for the "long" standard (contrary to what was observed in Experiment 1). A one-sample *t*-test with the averaged data from the 3.5 s and the 5 s samples revealed that the mean choice proportions of the comparisons associated with S1 and L2 ($M = 0.56$) was not significantly different from chance [$t(16) = 1.09$, $p = .15$, one-tailed].

Regarding the bisection points, a one-sample *t*-test showed that the bisection point of the psychometric function for Range 1 ($M = 3.67$; SE = 0.12) was not significantly different from the arithmetic mean of 3.5 s [$t(16) = 1.48$, $p = .08$, one-tailed]. Another one-sample *t*-test showed that the bisection point for Range 2 ($M = 4.88$; SE = 0.18) was not significantly different from the arithmetic mean of 5 s either [$t(16) = -0.66$, $p = .26$, one-tailed], contrary to what was observed in Experiment 1. Though the shifts in the bisection points with respect to the arithmetic means were in the direction predicted by the relative-coding hypothesis, they did not reach statistical significance.

The Weber ratios obtained for Range 1 and Range 2 were respectively equal to 0.16 (median = 0.19) and 0.18 (median = 0.11). A Wilcoxon signed rank test revealed no significant differences between either Weber ratio ($Z = -0.52$, *ns*). We conclude that temporal sensitivity for was similar for both ranges. Table 4 summarizes these results.

2.4.4.3. Comparing data from Experiment 1, Phase 4, and Experiment 2, Phase 5. We conducted two one-way mixed ANOVAS, one for Range 1 and the other for Range 2 with stimulus duration as within-subjects factor and experimental group as a between-subjects factor. The ANOVA on the data from Range 1 revealed a significant effect of stimulus duration [$F(4,128) = 199.65$, $p > .0001$], no significant group effect [$F(1,32) = 0.18$, *ns*], and no significant interaction effect [$F(4,13) = 0.14$, *ns*]. The ANOVA on the data from Range 2 revealed a significant effect of duration [$F(4,13) = 144.63$ $p < .0001$], no significant group effect [$F(1,32) = 0.22$, *ns*], and no interaction effect [$F(4,13) = 0.47$, *ns*]. These statistical results confirm the visual impression of a superposition of the psychometric functions obtained in the two experiments (Fig. 4).

A *t*-test for independent samples showed that the bisection point of the psychometric function for Range 1 in Experiment 1 ($M = 3.75$; SE = 0.15) did not differ significantly [$t(32) = 0.39$, $p = .70$, two-tailed] from the bisection point of the psychometric function for Range 1 in Experiment 2 ($M = 3.67$; S.E. = 0.12). Another *t*-test for independent samples showed that the bisection point for Range 2 in Experiment 1 ($M = 4.69$; S.E = 0.16) did not differ significantly [$t(32) = -0.78$ $p = .44$, two-tailed] from the bisection point of the psychometric function for Range 2 in Experiment 2 ($M = 4.88$; S.E. = 0.18).

A Mann–Whitney test revealed that the Weber ratio for Range 1 in Experiment 1 (median = 0.15) did not differ significantly ($U = 72.5$, *ns*) from the Weber ratio for Range 1 in Experiment 2 (median = 0.16). Another Mann–Whitney revealed no significant differences ($U = 85$, *ns*) between the Weber ratio for Range 2 in Experiment 1 (median = 0.17) and for Range 2 in Experiment 2 (median = 0.18). This suggests that the participants from both experiments had a similar sensitivity to temporal stimuli.

That the psychometric functions obtained in Experiment 1 did not differ from those obtained at the end of Experiment 2 suggests that behavior in the latter experiment was not influenced by the earlier, additional testing phase. Furthermore, given the absence of significant differences in terms of the proportion of "Long" responses, the location of the bisection points, or steepness of the slopes, we combined the psychometric functions from both experiments to increase the statistical power of our analyses.

2.4.4.4. Combining data from both Experiments. Fig. 5 shows the psychometric functions arising from the combined data. The mean proportion of "Long" responses at the arithmetic mean ($M = 0.432$) did not differ significantly from chance [$t(33) = -1.45$, $p = .079$, one-tailed] for Range 1, but it was significantly above chance [$t(33) = 1.87$, $p = .035$, one-tailed] for Range 2 ($M = 0.61$; SE = 0.06).

A one-sample *t*-test revealed that the mean bisection point of Range 1 ($M = 3.71$; SE = 0.09) was significantly above the

Table 4
Summary statistics for Experiment 2. *P* ("Long") at arithmetic mean is the mean proportion of "Long" responses when the sample duration equaled the arithmetic mean. L95% and U95% are respectively the lower and the upper limits of the 95% confidence interval for the mean.

Task	*P* ("Long") at arithmetic mean			Bisection points (s)			Slope of the psychometric function
	Mean	L95%	U95%	Mean	L95%	U95%	
Range 1	0.43	0.33	0.54	3.67	3.47	3.88	0.16
Range 2	0.56	0.43	0.70	4.88	4.56	5.20	0.18

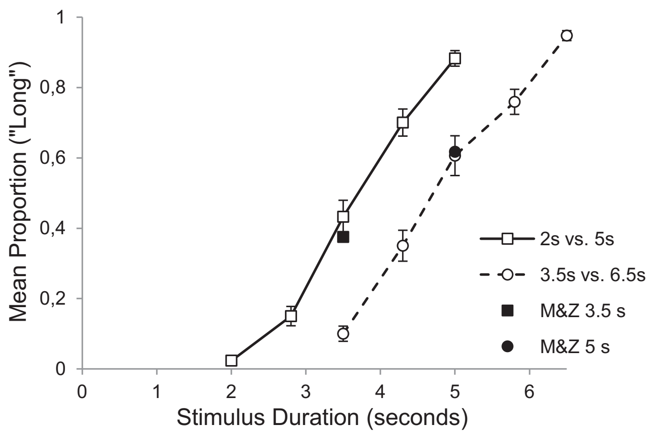

Fig. 5. Psychometric functions for the combined data (data from 34 subjects). Standard errors are displayed for each sample duration.

arithmetic mean of 3.5 s [$t(33) = 2.23$, $p = .016$, one-tailed]. Another one-sample t-test revealed that the mean bisection point of Range 2 ($M = 4.78$; SE = 0.12) was significantly below the arithmetic mean of 5 s [$t(33) = -1.78$, $p = .042$, one tailed].

By combining data from both experiments, we obtained a pattern of results closer to that reported by Molet and Zentall (2008). Both the proportion of “Long” responses at the arithmetic mean and the location of the bisection point suggest that the participants responded similarly to the “long” standards from each temporal range. Also, there was a slight indication for a grouping of the “short” standards because the bisection point for Range 1 was significantly above the arithmetic mean.

2.4.4.5. Testing in Phase 2. Testing in Phase 2 allowed us to obtain measures of behavior in the absence of any relative-coding effect induced by the second task. Thus, the obtained psychometric functions served as baseline for the test performed in Phase 5, giving us a better grasp of possible changes in the psychometric functions. Fig. 6 shows the psychometric functions obtained in Phase 2.

A repeated-measures ANOVA revealed that participants tended to give more “Long” responses as stimulus duration increased both for Range 1 [$F(4,32) = 43.58$, $p < .0001$] and for Range 2 [$F(4,28) = 48.39$, $p < .0001$]. The mean proportion of “Long” responses at 3.5 s ($M = 0.56$) and at 5 s ($M = 0.58$) were not significantly different from chance [$t(8) = 0.58$, $p = .58$, two-tailed, and $t(7) = 1.67$, $p = .14$, two-tailed, respectively]. The bisection points were observed at 3.25 s and 4.91 s for Range 1 and Range 2, respectively.

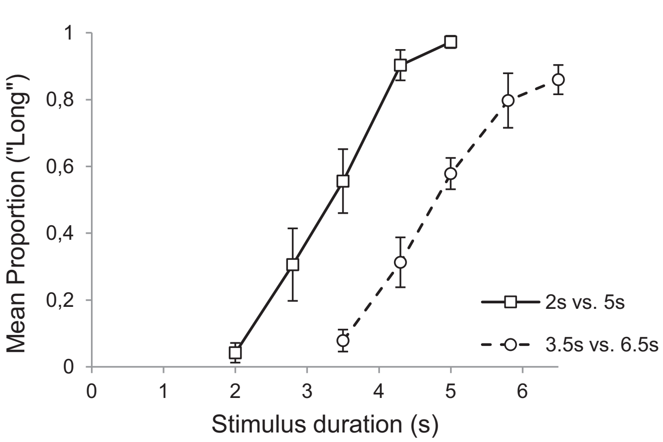

Fig. 6. Psychometric functions from Phase 2 of Experiment 2. Standard errors are displayed for each sample duration.

Extending the argument by Molet and Zentall (2008), relative coding implies that training on the second temporal task should induce a rightward shift in the psychometric function for Range 1 and a leftward shift in the psychometric function for Range 2, as compared to the psychometric functions obtained in Phase 2. To evaluate this hypothesis, we compared the psychometric functions obtained in Phases 2 and 5.

2.4.4.6. Comparing the data from Phase 2 and Phase 5. We conducted two repeated measures ANOVAS (within subjects factors: stimulus duration and test phase), one for Range 1 and the other for Range 2. One ANOVA revealed a significant [$F(4,32) = 110.33$, $p < .0001$] effect of duration, no significant effect of test phase [$F(1,8) = 3.93$, *ns*], and no significant interaction effect [$F(4,32) = 0.82$, *ns*] for Range 1. The other ANOVA revealed a significant effect of duration [$F(4,28) = 62.46$, $p < 0.0001$], no significant effect of test phase [$F(1,7) = 1.58$, *ns*], and no significant interaction effect [$F(4,28) = 0.88$, *ns*] for Range 2. These results demonstrate that there was no difference in the global pattern of the two psychometric functions obtained for each temporal range in two testing phases.

According to the relative-coding hypothesis as proposed by Molet and Zentall (2008), the proportion of “Long” responses at 3.5 s, which is both the arithmetic mean of Range 1 and the “short” standard in Range 2, should decrease in Phase 5 as compared do Phase 2. Also, the proportion of “Long” responses at 5 s, which is the arithmetic mean of Range 2 and the “long” standard in Range 1, should increase in Phase 5. We found that the proportion of “Long” responses at 3.5 s in Phase 2 ($M = 0.56$, SE = 0.1) and in Phase 5 ($M = 0.46$, SE = 0.1) were not significantly different from each other [$t(8) = -0.69$, $p = .26$, one tailed]. Another paired-samples t-test revealed no significant differences [$t(7) = 0.45$, $p = 33$, two tailed] between the proportion of “Long” responses for 5 s in Phase 2 ($M = 0.58$, SE = 0.05) and in Phase 5 ($M = 0.62$, SE = 0.13). Though the differences in the proportion of “Long” responses at the arithmetic means were not statistically significant across testing phases, they were in the direction predicted by the relative-coding hypothesis.

A paired-samples t-test showed that the bisection point for Range 1 in Phase 2 ($M = 3.25$; S.E = 0.19) was significantly smaller [$t(8) = -1.92$, $p = .04$, one-tailed] than the bisection point for Range 1 in Phase 5 ($M = 3.67$; S.E = 0.16). Such a downward shift in the bisection point is consistent with the relative-coding hypothesis (Molet and Zentall, 2008), and entails a shift of the psychometric function to the right from Phase 2 to Phase 5 (Fig. 7). Another paired-samples t-test revealed no significant difference [$t(7) = 0.59$, $p = .29$, one-tailed] between the bisection points for Range 2 obtained in Phase 2 ($M = 4.91$; S.E. = 0.19) and in Phase 5 (M = 4.79; S.E. = 0.30). The psychometric function for Range 2 did shift slightly to the left (Fig. 7), but the magnitude of the shift from Phase 2 to Phase 5 was smaller than that observed with Range 1.

A Wilcoxon signed rank test revealed no significant difference ($Z = -0.28$, *ns*) between the Weber ratios for Range 1 in Phase 2 (median = 0.11) and in Phase 5 (median = 0.17). Another Wilcoxon signed rank test revealed no significant difference ($Z = -0.41$, *ns*) between the Weber ratios for Range 2 in Phase 2 (median = 0.12) and in Phase 5 (median = 0.19). In sum, timing sensitivity for either stimulus ranges did not change significantly through training on the second task.

Again, our results do not provide clear evidence in favor of the relative-coding hypothesis. There was a significant change in the bisection point for Range 1 that was related to a decrease in the proportion of “Long” responses to the shorter sample durations in this range. However, analysis of the slopes of the psychometric functions did not confirm changes in timing sensitivity due to the changes in the bisection point. None of the other analyses

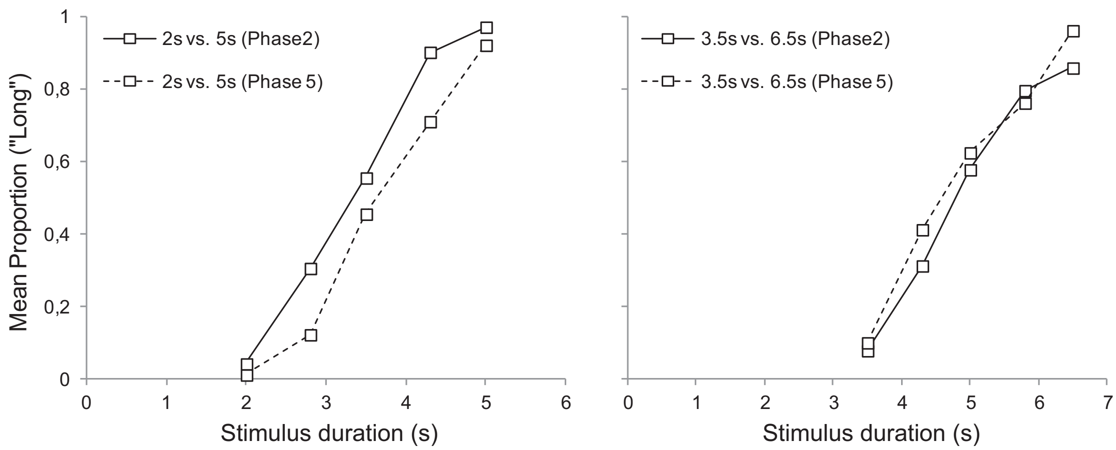

Fig. 7. Comparison of the psychometric functions obtained for each range in Phases 2 and 5. The left panel shows the data for Range 1. The right panel shows the data for Range 2.

yielded significant results, indicating that either the effect reported by Molet and Zentall (2008) is not robust, or that larger data sets are needed to detect effects of responding on the basis of the relative value of temporal stimuli.

3. General discussion

In the present study we evaluated the robustness of a finding reported with humans (Molet and Zentall, 2008) and pigeons (Zentall et al., 2004) and in which the subjects tend to respond similarly to stimuli that share the same relative value across temporal discrimination tasks. The authors of these studies suggested that subjects group temporal stimuli as "short" or "long" regardless of the absolute durations involved.

The critical test of this relative-coding hypothesis in humans is to use the arithmetic mean within a given range as a standard duration along with comparisons from another range (Fig. 1). It is well documented that human participants tend to bisect temporal intervals at the arithmetic mean of the trained standards (Wearden, 1991; Wearden and Ferrara, 1996; Wearden et al., 1997), and, therefore, tend to respond equally to each comparison stimulus when presented with a sample located at the arithmetic mean of its training range. According to Molet and Zentall (2008) and Zentall et al. (2004), participants who group "short" stimuli together as well as "long" stimulus together should choose comparisons previously associated with "short" following "short" samples and should choose comparisons previously associated with "long" following "long" samples. More specifically, when forced to choose between a pair of comparison stimuli, participants should choose the comparison previously associated with a "short" stimulus (S_1, for example) when presented with a "short" arithmetic mean (S_2), and should to choose the comparison associated with a "long" stimulus (L_2) when presented with a "long" arithmetic mean (L_1).

However, the relative-coding hypothesis stands on two data points, those corresponding to the arithmetic mean of each temporal task. The lack of data concerning the location of the bisection points and the shape of psychometric functions makes further evaluation of this hypothesis difficult. Here we replicated the basic procedure from Molet and Zentall's (2008) Experiment 2, but we also improved it by introducing several sample durations during testing, as is commonly done in bisection studies with humans (Allan and Gibbon, 1991; Wearden, 1991; Wearden and Ferrara, 1995, 1996).

The results from our experiments were unclear with regard to the relative-coding effect. The results from our Experiment 1 were mixed. On the one hand, within Range 1 neither the proportion of "Long" responses at the arithmetic mean nor the location of the bisection point provided statistically significant evidence in favor of a grouping of the "short" standards. On the other hand, both the proportion of "Long" responses at the arithmetic mean and the position of the bisection point within Range 2 supported the hypothesis of a grouping of the "long" standards. Furthermore, the results from Experiment 2 showed no significant relative-coding effect with respect to either the "short" or the "long" standards. However, after combining the data from both experiments, we found results compatible with a relational grouping of the "long" standards and some indication for a possible grouping of the "short" standards. Although our findings were not as robust as those reported by Molet and Zentall (2008), both the confidence intervals for the mean proportion of "Long" responses and the bisection points suggest that our participants' behavioral patterns were in the direction predicted by the relative-coding hypothesis (Tables 2 and 4).

The comparison of the psychometric functions obtained in Phases 2 and 5 of Experiment 2 was crucial for evaluating changes in the shape of the psychometric functions that might suggest grouping effects (Table 5). We found that the psychometric function for Range 1 was shifted to the right in Phase 5 (as compared to Phase 2) and that the psychometric function for Range 2 was shifted slightly to the left. However, these shifts were not of sufficient magnitude to translate into changes of temporal sensitivity, because the Weber ratios did not vary significantly between Phase 2 and Phase 5.

Overall, our results suggest that a change might occur in the pattern of responding after training on the second task. But because consistent positive evidence emerged only after combining data from two different Experiments, the detectability of the relative-coding effect may depend on increased sample size and reduced variability. Alternatively, our results (including the shifts of the psychometric functions from Phase 2 to Phase 5) may be explained in terms of alternative processes that do not rely on relative coding.

Consider the following proposition. When trained on a single bisection task, participants might learn to set a decision criterion that guides their choices of the comparisons associated with "short" or "long" standards. We can further assume that on the average, the criterion is centered on the arithmetic mean of the "short" and "long" standards. In Fig. 8 (panel I), hypothetical distributions of the criteria for a 2 s vs. 5 s discrimination and for a 3.5 s vs. 6.5 s discrimination are represented on the left and right panels, respectively. The criteria are centered over the hypothetical distributions of each temporal range and are represented in Fig. 8 (panel I) as vertical lines. Notice that one criterion passes over the arithmetic mean of Range 1 ($x = 3.5$) and the other over the arithmetic mean of Range 2 ($x = 5$). However, due to some intrinsic variability,

Table 5
Summary statistics of Phases 2 and 5 of Experiment 2. *P* ("Long") at arithmetic mean is the mean proportion of "Long" responses when the sample duration equaled the arithmetic mean. L95% and U95% are respectively the lower and the upper limits of the 95% confidence interval for the mean.

	Discr.	*P* ("Long") at arithmetic mean			Bisection points (s)		
		Mean	L95%	U95%	Mean	L95%	U95%
Range 1	Phase 2	0.56	0.38	0.73	3.25	2.89	4.10
	Phase 5	0.46	0.28	0.63	3.67	3.37	3.97
Range 2	Phase 2	0.58	0.49	0.67	4.91	4.54	5.27
	Phase 5	0.62	0.38	0.87	4.79	4.22	5.36

sometimes the criterion will fall above or below the arithmetic mean. The participant will choose the comparison associated with the "short" standard if the sample presented is smaller than the criterion, and will choose the comparison associated with the "long" standard if the sample is larger than the criterion.

When participants are trained on two different tasks, they may learn to set two different criteria, one appropriate for the first task and another appropriate for the second task (Fig. 8, panel II). However, each criterion is set in a different context. One is set in the context of the comparisons available in, say, Range 1 (context 1), and the other is set in the context of the comparisons trained in, say, Range 2 (context 2).

When a sample that was trained in one context (say, context 1) is presented in another context (say, context 2), the participant may be confused on a small percentage of trials (e.g. 10% of trials) and on those trials, instead of using the "correct" criterion acquired in context 1, uses the criterion acquired in context 2 (Fig. 8, panel II). Because the "correct" criterion will guide performance in less than 100% of the trials, specific shifts in the psychometric functions would be induced, depending on the "incorrect" criterion being above or below the "correct" one (Fig. 8, panel III). If below the "correct" criterion, there should be a leftward shift in the psychometric function (Fig. 8, panel III, right panel). If above, there should be a rightward shift (Fig. 8, panel III, left panel).

Consider how the explanation applies to our procedure. Participants trained on Range 1 should learn to respond 'red' following the 2 s standard and 'green' following the 5 s standard. This is so because the criterion in this task is set at the arithmetic mean (3.5 s), above 2 s and below 5 s (Fig. 8, panel I, left panel). Similarly, training on Range 2 should lead participants to choose the triangle

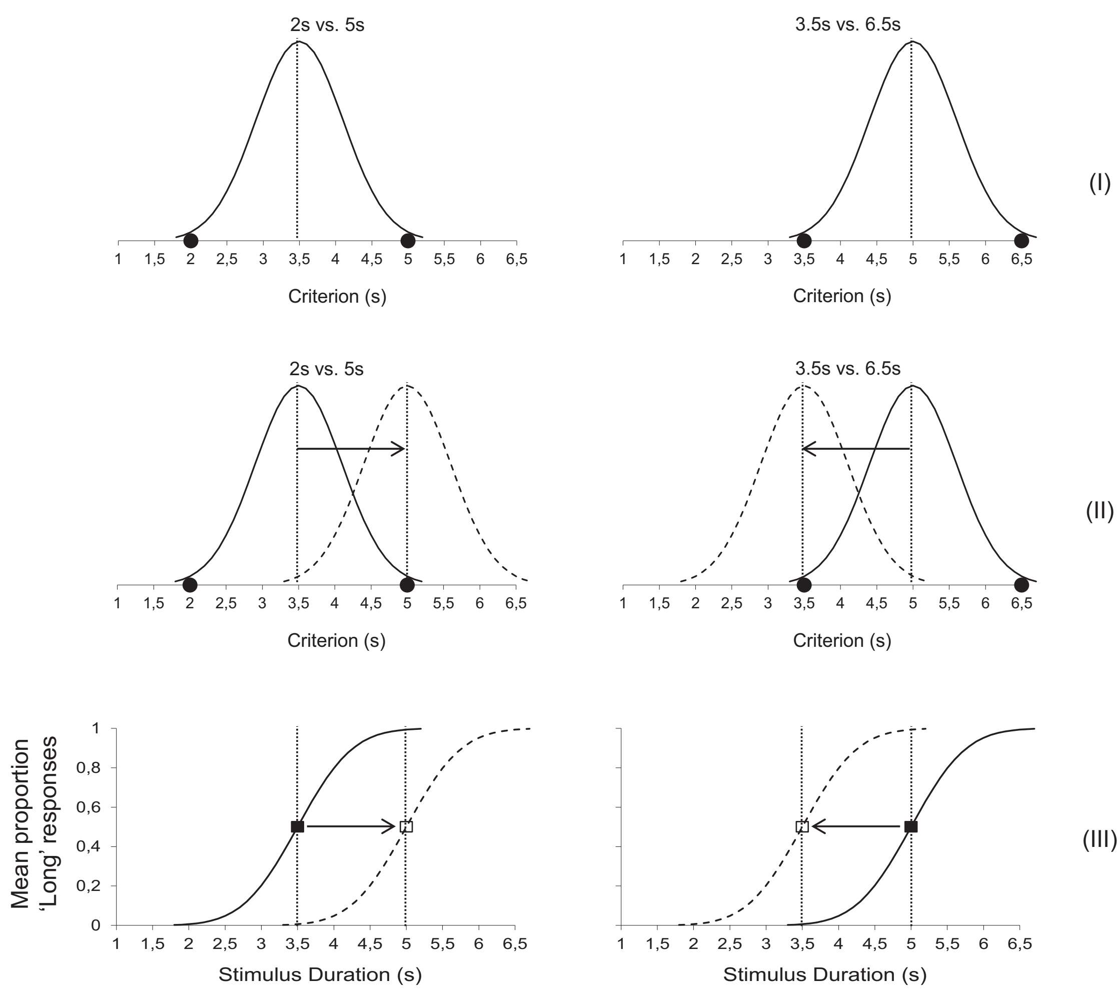

Fig. 8. Panel I shows hypothetical distributions of the criteria for a discrimination between 2 s vs. 5 s (left panel) and for a discrimination between 3.5 s vs. 6.5 s (right panel) following a single bisection training. Panel II shows the distribution of the criteria following a double bisection task. Panel III shows the corresponding psychometric functions for each discrimination task (black function) and the shifts in the psychometric functions induced by training in the second discrimination task (dashed function).

following the 3.5 s standard and the circle following the 6.5 s standard, because the criterion in this task is set at 5 s, above 3.5 s and below 6.5 s (Fig. 8, panel I, right panel).

When durations within Range 1 are tested in the context of red and green comparisons, participants may pick the 5 s criterion (developed for Range 2) on a small percentage of trials, because sample durations presented in testing Range 1 overlap with sample durations pertaining to Range 2. Because the "incorrect" criterion (5 s) is above the "correct" one (3.5 s) (Fig. 8, panel II, left panel), the participant would be biased toward the comparison associated with the "short" standard, that is, red. This bias would translate as a rightward shift of the psychometric function for Range 1 (Fig. 8, panel III, left panel).

When durations within Range 2 are tested in the context of triangle and circle comparisons, participants may mistakenly pick the 3.5 s criterion on a small percentage of trials because, again, durations within Range 1 overlap with the durations within Range 2. Because 3.5 s is below 5.0 s (Fig. 8, panel II, right panel), the participant would present a bias toward the option associated with the "long" standard. Hence we would observe a leftward shift in the psychometric function of Range 2 (Fig. 8, panel III, right panel).

The foregoing explanation accounts reasonably well for relative-coding-like data on the assumption of absolute coding of temporal intervals plus a probabilistic mixture of two criteria. Because it does not invoke higher or more complex processes, we argue that the relative-coding hypothesis is not needed to explain the pattern of results observed here and in other studies (Molet and Zentall, 2008; Zentall et al., 2004).

Acknowledgements

The work here reported is part of the Master thesis presented by the first author to the School of Psychology of the University of Minho. The work was supported by grants from the Portuguese Foundation for Science and Technology (FCT) to the co-authors.

References

Allan, L.G., Gibbon, J., 1991. Human bisection at the geometric mean. Learn. Motiv. 22, 39–58.
Carvalho, M.P., Machado, A., 2012. Relative versus absolute stimulus control in the temporal bisection task. J. Exp. Anal. Behav. 98 (1), 23–44.
Church, R., Deluty, M., 1977. Bisection of temporal intervals. J. Exp. Psychol. Anim. Behav. Process. 3, 216–228.
Hulse, S., Kline, C., 1993. The perception of time relations in auditory tempo discrimantion. Learn. Behav. 21, 281–288.
Maia, S., Machado, A., 2009. Representation of time intervals in a double bisection task: relative or absolute? Behav. Process. 81, 280–285.
Molet, M., Zentall, T., 2008. Relative judgments affect assessments of stimulus duration. Psychon. Bull. Rev. 15, 431–436.
Wearden, J.H., 1991. Human performance on an analogue of an interval bisection task. Q. J. Exp. Psychol. 43B, 59–81.
Wearden, J.H., Ferrara, A., 1995. Stimulus spacing effects in temporal bisection by humans. Q. J. Exp. Psychol. 48B, 289–310.
Wearden, J.H., Ferrara, A., 1996. Stimulus range effects in temporal bisection by humans. Q. J. Exp. Psychol. 49B, 24–44.
Wearden, J.H., Rogers, P., Thomas, R., 1997. Temporal bisection in humans with longer stimulus durations. Q. J. Exp. Psychol. 50B, 79–94.
Zentall, T., Weaver, J., Clement, T., 2004. Pigeons group time intervals according to their relative duration. Psychon. Bull. Rev. 11, 113–117.