Practice, stimulus complexity, and response bias on a target identification task.

Abstract

Four experiments assessed the effect of practice with similar stimuli on the ability of adult humans to identify a target stimulus by means of same/different judgments. The study found the target identification task sensitive to the effect of practice and to stimulus length and distinctiveness (Experiments 1 and 2). By the other hand, training improved performance on same but not in different trials (Experiment 3 and 4), being these findings apparently not related to any potential response bias. The target identification task was then sensitive to variables related to perceptual learning, providing some new insights about the potential biases to respond same or different. Further, it raised the hypothesis that the stimulus recognition and differentiation process might appear dissociated in such kind of task.

Keywords: Differentiation; Identification; Perceptual learning; Recognition; Response bias

Resumen

Cuatro experimentos valoraron el efecto de la práctica con estímulos similares en la habilidad de humanos adultos para identificar un estímulo muestra mediante juicios igual/diferente. El estudio encontró la tarea de identificación de la muestra sensible al efecto de la práctica, así como a la longitud y similitud de los estímulos (Experimentos 1 y 2). Además, se encontró que el entrenamiento mejoró la actuación de los participantes en la tarea en los ensayos de “igual” pero no en los de “diferente” (Experimentos 3 y 4), no estando estos resultados aparentemente relacionados con algún potencial sesgo de respuesta. En conclusion, la tarea resultó sensible a variables relacionadas con el aprendizaje perceptivo y aportó algo de luz sobre potenciales sesgos a responder igual o diferente. Además, levantó la hipotesis de que los procesos de reconocimiento y diferenciación de estímulos podrían aparecer disociados en esta tarea.

Palabras clave: Aprendizaje Perceptivo; Diferenciación; Identificación; Reconocimiento; Sesgo de Respuesta.

1. Introduction

Gibson and Gibson (1955) presented a target nonsense scribble for a few seconds to people of different ages, and then required them to identify it among a series of similar scribbles by means of same judgments. The study showed that younger children needed more repetitions of the scribbles to identify the target without errors compared with older children and adults. In addition, the number of errors was dependent on the number of dimensions on which the stimuli differed. More errors were made when the stimuli differed on one dimension than when they differed in terms of many features. In any case, people improved in their ability to identify the target with repeated presentations of the stimuli, although feedback was never provided. According to Gibson (1969), specificity for the "same" responses would have increased throughout repetitions of the stimuli because the ability to differentiate the target and non-target stimuli increased simply by repeated practice with the stimuli. This study has often been cited in the literature as the first to investigate progressive stimulus differentiation in the absence of reinforcement or feedback, i.e., perceptual learning (e.g., Hall, 2001). But tasks of the sort used by Gibson and Gibson (1955) were largely ignored in subsequent studies of perceptual learning (but see also Gibson, Gibson, Pick, & Osser, 1962).

Within the field of learning studies (but see also, Fahle & Poggio, 2002; Goldstone, 1998, for others from other theoretical approaches), the perceptual learning effect was then addressed mainly by studies conducted with non-human animals and conditioning preparations (e.g., Honey, Bateson & Horn, 1994, Honey & Hall, 1989; Mackintosh, Kaye, & Bennett, 1991, Symonds & Hall, 1995, 1997). Recently, however, interest in human perceptual learning appears to have been renewed (see for example Mitchell & Hall, 2014; Seitz & Dinse, 2007; Seitz & Watanabe, 2005). The ability of people to differentiate the stimuli after their pre-exposure has been tested in two main ways: by means of a categorization task with feedback (e.g., Mundy, Honey, & Dwyer, 2007, 2009), where participants must assign similar stimuli to different categories, or by tasks where similar stimuli presented successively a few seconds apart, must be judged as same or different (e.g. Dwyer, Hodder, & Honey, 2004; Lavis & Mitchell, 2006; Wang & Mitchell, 2011). Other kind of studies has assessed both performance and neuronal activity with other procedures and stimuli such as, for instance, those involved in studies about hyperacuity (i.e., Gilbert, Kapadia & Westheimer, 2000; Westheirmer & Gilbert , 1998; see also Ahissar & Hochstein, 1993; Crist, Kapadia, Westheimer & Gilbert, 1997; Shiu & Pashler, 1992). But only two studies have employed a target identification task similar to that used originally by Gibson (Angulo & Alonso, 2012; 2013). Those studies have yielded, however, some interesting results.

For example, Angulo and Alonso (2013) obtained evidence of an attentional shift in the processing of the stimuli—attention being directed toward the distinctive elements of similar stimuli and away from those that they shared in common, at least when pre-exposure schedule provided good opportunities to compare the stimuli. Previously, it had been found with the same target identification task that the optimal pre-exposure schedule for stimulus comparison (i.e., the concurrent pre-exposure schedule) increased the accuracy with which the participants were able to make a same/different judgement between the target and another similar stimulus (Angulo & Alonso, 2012). Nonetheless, this increase in accuracy on same/different judgement tasks was not accompanied by an increase in accuracy in recognizing the target in a multiple- choice task or reconstructing the target stimulus in a puzzle task. In brief, the findings reported by Angulo and Alonso (2012) seemed to show that pre-exposure schedule effects depend to some extent on the specific task used to measure such effects. Other authors have also suggested that encoding of the stimuli might be affected by the specific demands of the task (e.g, Gilbert, Kapadia & Westheimer, 2000; Seitz & Dinse, 2007; Westheirmer & Gilbert, 1998). To the extent that tasks might affect both stimulus encoding and the expression of the pre-exposure effects, the tasks employed to assess perceptual learning need to be more carefully analyzed in further studies since the particularities of the tasks might interact with the effects that they are designed to measure.

With this issue in mind, the principal aim of the present study was to explore the target identification task previously employed by Angulo and Alonso (2012, 2013) to assess pre-exposure schedule effects. This task was very similar to that originally employed by Gibson and Gibson (1955) but differed in terms of the visual stimuli used (nonsense Arabic character compounds instead of non-sense scribbles) and the number of different stimuli involved in the task. While in the Gibsonian task the target was presented among a series of scribbles that differed from the target, in the task designed by Angulo and Alonso the series of stimuli was composed of copies of the target and only one different stimulus. In order to establish how the target identification task behaves in a situation that more closely resembles the one originally proposed by Gibson, the present study required the participants to identify the target stimulus by same/different judgments among a group of 19 similar stimuli (Experiment 1) instead of a series of 20 stimulus presentations involving the target and only one other similar stimulus (Angulo & Alonso, 2012; 2013). Following this, the study explored how performance on the task is affected by important variables for perceptual learning such as the overall number of elements constituting the stimuli along with their number and proportion of distinctive and common elements (Experiment 2). Whilst the importance of the number of common elements of the stimuli for differentiation has already been extensively examined in the literature (see for example, Honey, Bateson & Horn, 1994; Mackintosh, Kaye & Bennett, 1991), the general effect of the number of elements on stimulus differentiation has been less well documented.

Finally, the present study also aimed to explore whether or not the task might be affected by a response bias towards judging the stimuli as same or different (Experiments 1-4). In general, the studies assessing stimulus differentiation by same/different judgments found greater accuracy when the correct response was to judge the stimuli as same rather than as different (e.g. Angulo & Alonso, 2012, 2013; Dwyer, Hodder, & Honey, 2004; Lavis & Mitchell, 2006, Wang & Mitchell, 2011). In fact, participants committed almost no errors when the correct response was "same", the effects of the pre-exposure schedules being detected only by the "different" correct response scores. On the basis of such evidence, it has been suggested that people might show a bias to respond "same" in tasks requiring "same-different" judgments (e.g., Lavis & Mitchell, 2006). One might think that "same" would be the expected response when the stimuli to be judged as same or different were similar and difficult to distinguish. Such a general tendency to respond “same” would then result in a greater percentage of errors on the trials involving different stimuli compared with those trials involving identical stimuli, and very few incorrect answers in the latter case. Given that the correct responses would be asymptotic on the “same trials” (hereafter “same trials” refers to those trials in which the correct response is same and “different trials” refers to those in which the correct response is different), the pre-exposure effects would be detectable only on “different trials”. As just described, a general trend to respond same might explain the findings that have emerged from the majority of those studies cited above. What remains to be clarified, however, is whether or not such a tendency to respond “same” should be considered a true response bias or an artifact caused by the similarity of the stimuli. If the latter were the case, the tendency to respond “same” should change over the course of the block of trials as the ability to differentiate the stimuli improves. A true response bias however, not should change with experience because it can be defined as a stable response. Furthermore, the trend to response “same” as an artifact based on the inability to differentiate the stimuli should be sensitive to variables affecting stimulus differentiation (for example, the pre-exposure schedule or stimulus complexity) but this would be not expected for a true and unconditioned response bias. With the exception of the studies reported by Angulo and Alonso (2012; 213), the accuracy on same/different judgments was presented as an average for all of the test trials (e.g. Dwyer, Hodder, & Honey, 2004; Lavis & Mitchell, 2006, Wang & Mitchell, 2011) and thus, it was not possible to see whether or not the hypothetical response trend might change with experience throughout the task or whether it would interact with other variables such as the pre-exposure schedule or stimulus similarity. Angulo and Alonso (2012; 2013) however, showed curves of accuracy for the “same” and “different” trials separately, finding some evidence that accuracy on these judgments might differentially vary across blocks of trials. Following concurrent pre-exposure (a pre-exposure schedule that particularly improves stimulus differentiation), for example, people began making more errors on different than on same trials but thereafter the errors on both types of trials were similar. This result might be taken to suggest that in this case, a stable response bias would not be operating and that the initial tendency to respond “same” disappears when there is an opportunity for the effects of stimulus differentiation to emerge. In addition to the above, Angulo and Alonso (2012) reported some evidence suggesting that the target identification task might be separately assessing two different processes. For the “same” judgments, the task might be assessing the ability of the participants to recognize the target stimulus whilst the “different” judgments might be assessing the ability to differentiate the other stimulus from the target (see Angulo & Alonso 2012, for further details). If this were the case, to test whether or not the task might be eliciting a response bias becomes particularly important because a tendency to respond “same” or “different” might have a different impact on the capacity of the task to detect the recognition and differentiation processes.

2. Experiment 1

Experiment 1 was conducted with the simple aim of testing whether a target identification task, very similar to that previously used by Gibson and Gibson (1955), could be sensitive to repeated practice with the stimuli chosen by Angulo and Alonso (2012; 2013). The stimuli were 20 arbitrarily chosen nonsense compounds of Arabic characters (see Figure 1). One of them was presented to the participants as the "target" for a few seconds before being required to identify it among a series of 19 similar stimuli by means of same/different judgments. Presentation of the target was followed by a series of 20 stimuli (1 exemplar of the target among 19 other similar cues). This procedure was repeated three times consecutively, resulting in three blocks of 20 trials, with the target always being presented at the beginning of each block of trials. Such parameters have been found to be effective in previous experiments assessing stimulus pre-exposure effects, and are able to detect progressive improvements in stimulus differentiation. If the task is sensitive to the effect of practice which enables the stimuli to be better differentiated, a progressive decrement in the percentage of errors would be expected across the blocks of trials. It should be noted that, as a result of the stimulus presentation schedule, in Experiment 1 the correct response was "different" in 19 of the 20 trials and "same" in only one of these. Thus, a general trend to respond "same" should lead to a high percentage of errors.

2.1. Method

2.1.1. Participants, apparatus, and stimuli

One hundred and twelve native Spanish (non-Arabic speaking) undergraduate students (age 18-30 years; mostly women, ratio 8:10) from the University of the Basque Country participated voluntarily in the experiment. All subjects gave their informed consent, were naïve to the exact problem being investigated by the experiment, and had never participated in similar experiments.

Twenty nonsense compounds of 5 Arabic characters were employed as stimuli (see Figure 1). Only one character was distinctive in each compound, the other four characters being common to both. Stimuli were presented on a computer monitor of a DELL-compatible PC, appearing in black over a white background.

2.1.2. Procedure

Experiment 1 was conducted collectively in a single session lasting 20 min. Firstly, the following instructions were displayed on the computer while they were read out loud by the experimenter: "Now, visual stimuli will appear on the screen. The first stimulus is called the target and you should observe it during the time it is present. The subsequent stimuli are named items. You have to indicate whether each of these stimuli is the same or different from the target. You will see the target and the other stimuli three times with a short rest period between presentations. Before the task begins, we will perform a brief training trial". This pre-training test was identical to the subsequent target identification task, but it involved very different stimuli (the scribbles used before by Gibson and Gibson, 1955) and considerably fewer trials (4 - the "same" response being correct on half of such trials and "different" being correct on the others). After the pre-training phase, the task began with the presentation of a white screen with the word "target" in the centre for 3 s, indicating the onset of the target. The target stimulus was always the Arabic character compound labeled with the number 11 in Figure 1, and it was presented for 5 s. Then, all participants received a set of 20 trials consisting of a single stimulus presentation for 5 s in the centre of the screen with an interval between presentations of 3 s. During this interval, a white screen indicating the presentation of the following item was displayed. A different stimulus was presented on each trial and only one was the exact target. This procedure was repeated three times consecutively, with an interval of 10 s between repetitions, comprising a total of three blocks of 20 trials each. The target stimulus was the same on the three blocks of trials, but on each block it was placed in a different position in the series (position 11, 5, and 15 in the first, second, and third block, respectively). For the remaining stimuli, the position in the series was changed randomly on each block of trials. The responses of the participants were collected in written form and they did not receive feedback about their accuracy.

The dependent variable in this and subsequent experiments was the percentage of errors committed on each block of trials. Data were evaluated by analysis of variance (ANOVA) adopting a significance level of $p < .05$.

2.2. Results and Discussion

The Mean percentage of errors made by the participants in Experiment 1 were 30.04 (SEM±1.26), 25.31 (SEM±1.28) and 23.25 (SEM±0.97), in the first, second, and third block of trials, respectively. The percentage of errors was very small on the first block of trials but in spite of this, it appeared to decrease across blocks. An ANOVA conducted on the errors found this decrease to be significant, $F(2, 222) = 20.64$, $p < 0.001$.

This finding suggests that participants improved in their ability to differentiate the stimuli across blocks of trials, with their "different" responses increasing in specificity. The percentage of errors decreased even though it was very small at the beginning of the task. Thus, it seems that the task was quite sensitive to the effect of practice with the stimuli. At the same time, the small percentage of errors found in Experiment 1 provided little scope to support the idea of a general trend to respond "same" in this case. Because the correct response was "different" on 19 trials and "same" on only one of them (for each block of trials), one might suppose that a bias to respond "same" would result in a greater percentage of errors. This was not the result found. But it is possible that the stimuli were easily differentiated from the beginning of the task, in which case the hypothetical trend to respond same would then be counteracted by the ability to differentiate the target stimulus from the others. In this case, if the stimuli were more complex and similar, the initial percentage of "same" responses (and thus, errors) would be greater. Experiment 2 was conducted with the aim of testing whether the task was sensitive to the complexity of the stimulus, and in particular whether the use of complex stimuli would lead to a greater percentage of “same” responses.

3. Experiment 2

Experiment 2 assessed the effect of stimulus complexity on the task by manipulating two variables: the number of elements constituting the stimuli as a whole (namely, stimulus length), and the number of common elements shared by the stimuli (distinctiveness). So, for half of the participants in Experiment 2 the stimuli presented in the task were short (5 elements, groups S-Hi and S-Lo) while for the other half the stimuli presented were long (10 elements, groups L-Hi and L-Lo). Further, for half of the participants of the previous conditions the stimuli were of high distinctiveness (one element common to the stimuli, with all others being unique to each stimulus - Groups L-Hi and S-Hi) while for the other half the stimuli were of low distinctiveness (one element distinctive or unique to each stimulus, with all other elements being common to the stimuli - Groups L-Lo and S-Lo). Given the widely accepted notion that difficulty in differentiating between stimuli relies on the amount of common elements they share, (differentiation being harder between stimuli that share more features), more errors would be expected with the stimuli of low distinctiveness than with the stimuli of high distinctiveness. However, it might also be thought that the proportion of common elements could be at least as important as the overall amount of common elements in terms of differentiating the stimuli. One distinctive element among ten elements would render the stimuli more similar than one among five. Similarly, one common element among five should make the stimuli more similar than one among ten. Thus, more errors would be expected with the long than the short stimuli when the stimuli are of low distinctiveness and the opposite when the stimuli to be discriminated are of high distinctiveness.

3.1. Method

3.1.1. Participants, apparatus, and stimuli

Forty-five native Spanish (non-Arabic speaking) undergraduate students (age 18-30 years; mostly women, ratio 8:10) from the University of the Basque Country participated voluntarily in the experiment. All subjects gave their informed consent, were naïve to the exact problem being investigated by the experiment, and had never participated in similar experiments.

Seventy-eight compounds of Arabic characters were employed as stimuli. The short-low distinctiveness stimuli were exactly the same as those employed in Experiment 1 (see Figure 1). The stimuli long-low distinctiveness were formed by adding five common elements (always the same) to the short-low distinctiveness stimuli (for examples, see Figure 2). The stimuli short-high distinctiveness shared one element, the other four elements being different in each stimulus (see Figure 2). Finally, the stimuli long-high distinctiveness were formed by adding another five elements, always different, to the short-high distinctiveness stimuli (for examples, see Figure 2). All the details not specified here were identical to those described for Experiment 1.

3.1.2. Procedure

Participants were randomly assigned to four groups S-Lo (n= 11), S-Hi (n=11), L-Lo (n=10) and L-Hi (n=13). Groups differed only in the stimuli presented on the task, these being the short-low distinctiveness, short-high distinctiveness, long- low distinctiveness and long- high distinctiveness, in the groups S-Lo, S-Hi, L-Lo and L-Hi, respectively. The target stimulus was the same as that in Experiment 1 for the participants receiving the short stimuli (groups S-Lo, S-Hi). For the participants that received the long stimuli, the target was formed by adding five elements (the same five elements common to all of the long-low distinctive stimuli) to the target used in Experiment 1. Different to Experiment 1, here and in the following experiments the task was run individually on personal computers. This procedural change was introduced to avoid the potential effects of distance to the screen or angle of vision that might increase the variability in the responses and hinder the probability of detecting the complexity effects. In all other details not specified here, the experiment was conducted in exactly the same way as Experiment 1.

3.2. Results and Discussion

The percentage of errors across the three blocks of trials for the four groups can be seen in Figure 3. It appears that the percentage of errors was greater with the stimuli of low distinctiveness than high distinctiveness and also greater with the long than with the short stimuli. In any case, the percentage of errors seemed to decrease across blocks of trials. A 2 x 2 x 3 ANOVA with Stimulus length, Distinctiveness, and Block of trials was conducted on the data represented in Figure 3. This analysis revealed significant main effects of Length, $F(1, 41) = 8.81$, $p = 0.005$, Distinctiveness, $F(1, 41) = 59.33$, $p < 0.001$, and Block, $F(2, 82) = 12.16$, $p < 0.001$. No significant interactions were found between the variables, $F\text{s} \leq 1.06$.

Due to the fact that more errors were made with the low distinctiveness stimuli than with the high distinctiveness stimuli, the results supported the general assumption that the number of common elements is an important factor in discriminating the stimuli. Interestingly, the experiment failed to find an interaction between distinctiveness and length, the errors always being greater with the long than with the short stimuli. In the light of this finding, it might be concluded that the proportion of common elements was not as important as the overall number of common elements. But of course, it could also be possible that the task was not sensitive enough to detect such an effect. Irrespective of the distinctiveness of the stimuli, participants made more errors with the long than the short stimuli. At least to the best of our knowledge, the effect of the full number of elements present in the stimulus has yet to be tested in human perceptual learning studies. But in this task, the general effect of stimulus length might easily be explained in terms of memory. In order to judge each stimulus as being same or different to the target, participants had to maintain in memory the trace of the stimulus. So, the ease with which the target can be remembered on each trial will affect the accuracy of same/different judgments. According to general associative theories of learning (e.g., Rescorla & Wagner, 1972, Wagner, 1981), a memory for the stimulus would be progressively built by the establishment of excitatory links between the elements constituting the stimulus, i.e., by the unitization process (see for example, McLaren, Kaye, & Mackintosh, 1989; McLaren & Mackintosh, 2000). Thus, an accurate memory of the long stimuli would require more experience than for the shorter stimuli because there would be more elements to be linked in the former case than in the latter. Furthermore, the memory trace of the stimuli would be expected to decay after stimulus presentation, with progressively more and more details being lost. Clearly, the longer stimuli would be containing more details to lose than the shorter. Thus the detrimental effect of time might be greater for the longer than the shorter stimuli, the probabilities of failing to accurately identify the target being greater with the former case. From a different point of view, if the stimuli were being processed in a supervised way (induced by the instructions for the task) rather than an unsupervised way (for discussion of this issue, see, for example, Nelson, 2009), the greater length of the stimuli as well as the higher number of elements might have had a greater attentional cost, hindering the stimulus processing and encoding.

As found in Experiment 1, the initial percentage of errors in Experiment 2 was very small (no more than 45% in the more difficult condition). And given that the “different” response was correct in 19 of the total 20 trials in each block, a percentage of errors below 50% hardly provides support for the idea of a general bias to respond “same” and, if anything, the results raised the possibility that the participants might exhibit a trend to respond “different”. Whilst previous studies have reported evidence for just the opposite, Experiments 1 and 2 differed from these studies in at least one important aspect that might affect the hypothetical response bias - the variability of the stimuli presented. Previously, Angulo and Alonso presented only two different stimuli in the task whereas 20 stimuli were presented here. Thus, it might be that a greater number of different stimuli would lead to a greater number of “different” responses. Of course, it might be also possible that the stimuli used in Experiment 1 and 2 were very easily differentiated from the target, and the supposed tendency to respond “same” was then counteracted by the ability to differentiate the stimuli. But this possibility contrasted strongly with the verbal reports of participants at the end of the experiment. The task was described as very difficult, especially with the long stimuli of low-distinctiveness, and none of the participants reported that only a target stimulus was included in each block of trials. It is possible that the stimuli might be easily differentiated, in spite of the subjective impressions of the participants. Leaving aside this point for the moment, the principal aim of Experiment 3 was to test directly the percentage of “same” and “different” responses made by the participants in the task, by matching the trials on which the correct responses were “same” or “different”.

4. Experiment 3

Experiment 3 was conducted with the aim of testing whether participants could be displaying a response bias in the task. Previous studies have found consistently greater accuracy to judge two stimuli as “same” than as “different” (e.g. Dwyer, Hodder, & Honey, 2004; Lavis & Mitchell, 2006, Wang & Mitchell, 2011), the suggestion being that participants would be showing a trend or bias to respond “same” (see for example, Lavis and Mitchell, 2006). In order to test this possibility, the number of “same” and “different” correct responses was matched on each block of trials. If participants were displaying a bias to respond “different”, one would to expect more errors on the same trials (in which the stimulus presented was the target) than on different trials (in which the stimulus presented was another, different to the target). If the participants were displaying a bias to respond “same”, one would expect just the opposite result. Furthermore, if the participants exhibit an initial trend to respond “same”, and this is counteracted by stimulus differentiation, the initial percentage of errors should be greater for the “different” than the “same” trials. But across a block of trials, the percentage of “same” responses should decrease, leading to a decrement in the errors committed on the different trials. At the same time, increasing the number of presentations of the target might allow for assessment of both potential improvements in the ability to recognize and identify the target as well as differentiation of the target from the other stimuli, as in the previous experiments. In order to maintain the difficulty of the stimulus differentiation in the task, in Experiment 3 we used the short and long low distinctiveness stimuli that were more frequently confused with the target in Experiment 2.

4.1. Method

4.1.1. Participants, apparatus and stimuli

Thirty-eight (non-Arabic speaking) undergraduate students (age 18-25 years; mostly women, ratio 8:10) from the University of the Basque Country participated voluntarily in the experiment. All subjects gave their informed consent, were naïve to the exact problem being investigated, and had never participated in similar experiments.

In this experiment the stimuli employed were 22 compounds of Arabic characters, short low distinctiveness for half the participants, and long low distinctiveness for the remainder. In particular, the short low distinctiveness stimuli used in the task were those labelled as 1, 2, 3, 4, 7, 8, 10, 15, 16, and 17 in Figure 1, the long low distinctiveness stimuli simply being the long version of these (see the examples in Figure 2). The target stimuli were those presented for the groups S-Lo and L-Lo in Experiment 2, i.e., the stimulus labelled as 11 in Figure 1, and its long version, respectively.

4.1.2. Procedure

Participants were randomly assigned to two equal groups (groups S-Lo and Lo; n= 19). The groups differed only in the stimuli presented on the task, short low distinctiveness being presented for Group S-Lo, and long low distinctiveness for Group L-Lo. For all participants, blocks of trials were comprised of 10 stimuli identical to the target and a further 10 different (and different from each other) stimuli presented in alternation. In all other details not specified here, the experiment was conducted in exactly the same way as Experiment 2.

4.2. Results and Discussion

Performance on the task for the two groups is displayed in Figure 4. It appears that more errors on “different” trials than “same” trials were made in general, the latter decreasing (but not the former) across blocks of trials. Groups seemed to differ on the “different” trials but not on the “same” trials, the percentage of “different errors” being greater with the long than with the short stimuli. A 2 x 2 x 3 ANOVA with Group, Trial (same or different) and Block conducted on the data represented in Figure 4 found significant main effects of Group, $F(1, 36) = 16.92$, $p < 0.001$, Trial, $F(1, 36) = 5.07$, $p = 0.03$, and Block, $F(2, 72) = 17.89$, $p = < 0.001$. The double interactions, Group x Trial, $F(1, 36) = 3.78$, $p = 0.060$, and Trial x Block, $F(2, 72) = 2.72$, $p = 0.072$, were not significant (all Fs < 1). In spite of such interactions not reaching the criterion of statistical significance, subsequent analysis of simple effects were conducted in order to elucidate whether this task could be replicating some effects previously found with other tasks involving same/different judgments. Specifically, we wanted to test whether or not participants made more errors on “different” trials than on “same “trials, the effect of stimulus length being detected on “different” but not on “same” trials (as in previous studies where pre-exposure schedule effects were detected in the “different” trials but not on the “same” trials, e.g., Lavis & Mitchell, 2006; Mitchell, Nash & Hall, 2008).

Subsequent analysis found that participants made more errors on the “different” trials with the long than with the short stimuli $F(1, 36) = 19.15$, $p < 0.001$, while errors on “same” trials were similar, $F(1, 36) = 0.08$, $p = 0.774$. Furthermore, only with the long stimuli were the errors greater on the “different” than on the “same” trials, $F(1, 18) = 7.34$, $p = 0.014$. Errors on “same” trials, $F(2, 74) = 8.95$, $p < 0.001$, but not on “different” trials, $F(2, 74) = 0.779$, $p = 0.463$, decreased across blocks of trials, the errors being fewer on “same” trials than on different trials only in the last block of trials, $F(1, 37) = 11.12$, p = 0.002, remaining blocks, $Fs(1, 37) \leq 2.46$, ps ≥ 0.125.

In brief, participants made more “same” than “different” responses in general, leading to fewer errors on the same than on different trials. This result precludes the possibility that participants could be showing a bias to respond “different”. But it is not clear whether they might be showing a bias to respond “same”. If participants were showing a general trend to respond “same” *a priori*, one would expect there to be more errors on different than on same trials, not only on the last block of trials but from the beginning of the task, and not only with the long stimuli but also with the short. In addition, the reduction of errors on the same trials could not be explained solely on the basis of an enhancement of indiscriminate “same” responses across blocks of trials. In this latter case, errors on different trials should have increased while errors on the same trials decreased. But this is not the case for the results found in Experiment 3. It appeared that the number of “same” responses increased across blocks of trials in a discriminative way, these responses being confined only to the stimuli that were actually the target. In accord with Gibsonian theory (Gibson, 1969), this increment in the specificity for the same responses might be taken to indicate that perceptual learning was occurring. This raises the question then, as to why specificity for the different responses did not increase in the same way, as well as why the effect of stimulus length was detected by "different" trials but not by "same" trials.

It should be noted that, in order to match the number of trials in which the correct response was "same" and "different", the target stimulus was presented 10 times per block whilst the other 10 different stimuli were presented only once in each block. Thus, differences in the specificity for the "same" and "different" responses could have been affected by this unequal experience with the target and the other stimuli. Before entering into further discussion on this issue, Experiment 4 was conducted to empirically test this possibility. If the different performance for the "same" and "different" trials in Experiment 3 was because the experience with the stimulus involved on the "same" trials was greater than with those involved on the "different" trials, such a difference in performance should disappear when experience with the target and the other different stimulus is matched.

5. Experiment 4

In order to test the effect of the amount of target and non-target stimulus presentations on the results described above, two new groups were added to those used in Experiment 3. In these groups (S-Lo/2 and L-Lo/2) only two stimuli, the target and other non-target similar one, were presented in alternation 10 times each. The stimulus selected as our non-target was the one most frequently confused with the target in Experiment 2.

5.1. Method

5.1.1. Participants, apparatus, and stimuli

Forty-four (non-Arabic speaking) undergraduate students (age 17-35 years; mostly women, ratio 8:10) from the University of the Basque Country participated voluntarily in the experiment. All participants gave their informed consent, were naïve to the exact problem being investigated, and had never participated in similar experiments. The stimuli employed here were the same 22 compounds of Arabic characters employed in Experiment 3.

5.1.2. Procedure

Participants were randomly assigned to four equal groups (groups S-Lo, L-Lo, S-Lo/2 and L-Lo/2; n= 11). Groups differed only in terms of the stimuli presented on the task, these being short low distinctiveness for Group S-Lo and S-Lo/2, and long low distinctiveness for groups L-Lo and L-Lo/2. For groups S-Lo and L-Lo, the procedure was exactly the same as that described in Experiment 3. Thus, Experiment 4 can be considered a replication of such experimental conditions. The new groups received presentations of only two stimuli - the target, labelled with the number 11 in Figure 1, and the non-target stimulus labelled with the number 2 in Figure 1 (and their long version for the group L-Lo/2). Such stimuli were always presented in an intermixed schedule since it is well established that the stimulus presentation schedule has an effect on stimulus differentiation, and using always exactly the same schedule might serve to control for any effect produced by the aleatory presentation of the stimuli. Any effect of this kind, even being aleatory, might blur the results. Thus, for all the participants, "same" and "different" trials were always presented in alternation. Because in all other details not specified here, the experiment was conducted in exactly the same way as Experiment 3.

5.2. Results and Discussion

The Mean percentage of errors on same and different trials for the four groups of Experiment 4 can be seen in Figure 5. It appears that errors on the "same" trials were more markedly decreased than on "different" trials across blocks of trials. In general, the percentage of errors was greater on "different" than on "same" trials and only in the former case did the groups clearly differ. The errors on "Different" trials seemed to be greater when only two stimuli were presented than when the series included ten different stimuli, the errors also being greater with the long than the short stimuli. A 2 x 2 x 2 x 3 ANOVA with Stimulus length, Stimuli (2 or 20), Trial, and Block as the factors revealed the four main effects to be significant, Length, $F(1, 40) = 7.11$, $p = 0.011$, Stimuli, $F(1, 40) = 13.19$, $p = 0.001$, Trial, $F(1, 40) = 51.23$, $p < 0.001$, and Block, $F(2, 80) = 6.22$, $p = 0.003$. The double interaction Stimuli x Trial was significant $F(1, 40) = 8.65$, $p = 0.005$, whilst also approaching significance was the Length x Stimuli x Block interaction, $F(2,80) = 3.07$, $p = 0.052$; Length x Trial x Block, $F(2, 80) = 2.66$, $p = 0.076$, all remaining Fs ≤ 2.11. Subsequent analyses revealed the following. Participants made more errors on the "different" trials when the series involved 2 stimuli than when it involved 10, $F(1, 42) = 16.26$, $p < 0.001$, but in both cases participants made more errors on "different" than on "same" trials, with 10 stimuli, $F(1, 21) = 8.48$, $p = 0.008$, and with 2 stimuli, $F(1, 21) = 50.72$, $p < 0.001$. The triple interaction between Length x Stimuli x Block might be attributable to the fact that in the first blocks of trials, stimulus Length had an effect only when the task involved 11 stimuli, $F(1, 20) = 8.14$, $p < 0.001$. In the latter case the percentage of errors with the Short stimuli was 31.81%, and 50% with the Long Stimuli, while the percentage of errors was 50% and 50,90% with the Long and Short stimuli respectively, when only 2 stimuli were presented on the task. Finally, although the triple interaction Length x Trial x Block not reached the statistical significance, some analysis of simple effects were conducted in order to test whether the principal findings of Experiment 3 were replicated here. Effectively, this analysis confirmed that errors on "same" trials, $F(2,84) = 6.88$, $p = 0.002$, but not on "different" trials, decreased across blocks, and that the effect of stimulus Length was significant on "different" trials, $F(1, 42) = 4.73$, $p = 0.035$, but not on the "same" trials.

Experiment 4 confirmed that improvements in identifying the target by "same" judgments could appear without evidence of an equivalent improvement in stimulus differentiation by "different" judgments. Again, stimulus length affected the accuracy on "different" but not on the "same" trials. But increasing the number of presentations of a unique non-target stimulus appeared not to improve performance on "different" trials, but rather the reverse. The greater percentage of errors on the "different" trials when the task involved only 2 different stimuli compared to when the task used 11, is not surprising given that the stimulus chosen as the non-target stimulus in the first case was precisely the one more frequently confused with the target in Experiment 2. In any case, this result suggests that the improvement on "same" but not on "different" trials found in Experiment 3 is not due to the fact that the target was presented more extensively than the other stimuli.

6. General Discussion

The present study aimed to explore performance in a task very similar to that originally designed by Gibson and Gibson (1955) to assess improvements in stimulus differentiation as consequence of non-reinforced exposure to the stimuli (i.e, perceptual learning). Currently, there is growing evidence for the hypothesis that the specific tasks employed to assess human perceptual learning might determine how perceptual learning is expressed (i.e., Angulo & Alonso, 2012), as well how the stimuli might be processed (e.g, Gilbert, Kapadia & Westheimer, 2000; Seitz and Dinse, 2007; Westheirmer & Gilbert, 1998). Therefore, elucidating the intrinsic effects generated by a task have become of critical importance to understand the effects of such variables on perceptual learning. In particular, the task studied here was previously found to be sensitive to pre-exposure schedule effects (Angulo and Alonso, 2012; 2003), such effects being detected by the accuracy on “different” trials but not on “same” trials.

The present study appears to show that a target identification task involving the stimuli designed by Angulo and Alonso (2012, 2013) might be very sensitive to relevant variables involved in perceptual learning. Even when the initial percentage of errors was small (Experiment 1 and 2), such errors were significantly reduced across blocks of trials, indicating that the task is very sensitive to the effects of practice. When the stimuli presented in the task were long (10 elements), participants made more identification errors than when the stimuli were short (5 elements), and errors were also more evident when the stimuli shared many of their elements in common (4/5 or 9/10) than only one (1/5 or 1/9). Thus, the task was able to detect the effect of variables related to the ease with which the stimuli can be discriminated, such as stimulus length and distinctiveness (Experiment 2). Some previous studies have established the importance of the number of common elements of the stimuli for stimulus differentiation. There was no evidence, however, for the possibility that the overall amount of stimulus elements might also affect perceptual learning. Experiment 2 showed that, importantly, stimulus differentiation might be hindered by increasing the number of stimulus elements even when the added elements might be in fact decreasing stimulus similarity by the increment in the number of distinctive elements. Finally, the initial percentage of errors reached the 80% level in the last experiment only when two stimuli of low distinctiveness were presented in the task, although errors also decreased across blocks of trials (Experiment 4). Thus, the stimuli chosen for the task by Angulo and Alonso (2012, 2013) would be difficult to discriminate at the start of training, although stimulus differentiation would be potentially improved by practice and by previous experience with the stimuli. This appears then, to be an optimal protocol for assessing perceptual learning.

In addition, the study of the potential response bias could provide some interesting insights about the meaning of the “same” and “different” responses in the task. As with other tasks assessing stimulus differentiation by means of same/different judgments (e.g., Dwyer, Hodder & Honey, 2004; Lavis & Mitchell, 2006, Wang & Mitchell, 2011), here participants also seemed to generally make more errors on the “different” trials than on the “same” trials. Given that “same” would be the expected response when the stimuli cannot be distinguished, this general finding is not surprising. Participants would be responding “same” in an indiscriminate way, failing on “different” trials but being correct on the “same” trials. The small percentage of errors found in Experiments 1 and 2 clearly challenge the idea that the task itself might be activating an automatic and unconditioned bias to respond “same”. Because “different” would be the correct response in 19 of the 20 trials of each block, a bias to respond “same” would result in a greater percentage of errors. The findings yielded by Experiments 3 and 4 could not also be fully explained on the basis of a simple response bias to respond “same”. In Experiment 3, differences between same and different errors did not appear from the beginning of the task, as would be expected if participants were showing a general trend to respond “same”. Such differences emerged across blocks of trials because errors on the “same” but not on the “different” trials decreased. Neither could the errors on the “same” trials be explained by a progressive trend to respond “same” indiscriminately across blocks of trials, since if this were the case, then errors on “different” trials should have increased while errors on the “same” trials decreased. It is true that this latter seemed to be occurring in Experiment 4 when the task involved only two long stimuli. However, given that this the most difficult condition, and the same evidence was not found in the remaining conditions (see also the previous published works of Angulo and Alonso, 2012, 2013), such a result might be better explained as consequence of demotivation. Because the task was very difficult with such stimuli, after two blocks of trials without finding the differences between the stimuli participants might always choose to respond “same”.

The differences between “same” and “different” trials were noticeably greater with the long than with the short stimuli, and the effects of variables presumably related to discrimination difficulty, such as stimulus length and the stimuli involved in the task, were detected on “different” but not on “same” trials. This last result is also consistent with other studies where the effect of the variables tested is detected by the trials where the correct response is “different”. Finally, the results of Experiments 3 and 4 seemed to suggest that the different performance shown by the participants between the “same” and “different” trials was not caused by differences in the amount of correct “same” and “different” trials, or by differences in experience with the stimuli involved in the “same” and “different” trials.

The results of Experiments 3 and 4, therefore, seem to indicate that the participants improved in their ability to identify the target stimulus by “same” judgments to a greater extent than “different” judgments, their judgements being more accurate in general in the former case. If it is accepted that in this task, the correct “different” responses implies that participants were able to differentiate the target stimulus from the other similar stimuli, while the correct “same” responses would indicate that they were able to recognize the target, one might think that here, the stimulus recognition and differentiation processes are dissociated. Experience with the stimuli would enhance stimulus recognition to a greater extent and before than stimulus differentiation, and the variables related to the difficulty in differentiating the stimuli, such as stimulus length, would have a stronger impact on the differentiation process than the recognition process. The notion, *a priori,* of a dissociation between stimulus recognition and stimulus differentiation seems counterintuitive and, one might suppose that stimulus recognition implies stimulus differentiation and vice versa. In addition, most of the evidence supporting this hypothesis in the present study emerged from the analysis of no significant interactions. But whatever the case, the present findings are entirely consistent with others reported previously (Angulo

& Alonso, 2012). In one experiment, participants received concurrent, intermixed or blocked pre-exposures to the stimuli before testing the effects of these pre-exposure schedules on different tasks. Participants receiving concurrent pre-exposures to the stimuli were more accurate than the others in identifying the target stimulus by means of different judgments in a task identical to that used in Experiment 4 with the short stimuli. When participants were then required to identify the same target stimulus in a multiple choice task, participants who received the concurrent and intermixed pre-exposure schedules confused the two pre-exposed stimuli to a lesser extent than those who had received the blocked schedules. As with other findings reported previously (Mundy, Honey & Dwyer, 2007, 2009), the latter also seems to indicate that participants were more accurate in differentiating between the target and non-target stimuli after concurrent or intermixed, than with blocked pre-exposures to the stimuli. However, all the participants seemed to be similarly accurate in identifying the target stimulus by means of “same” judgments on the target identification task, as well as by their selections on the multiple-choice task. Thus, irrespective of the pre-exposure schedule received, all the participants seemed to be similarly accurate in recognizing the target stimulus.

Current accounts of perceptual learning (e.g., Hall, 2003; McLaren, Kaye, & Mackintosh, 1989; McLaren & Mackintosh, 2000; Mitchell and Hall, 2014; Mitchell, Nash et al., 2008) do not explicitly recognize the possibility that stimulus recognition and differentiation can be dissociated in some circumstances, their explanatory scope being limited to the case of stimulus differentiation. But our findings suggest that this possibility should at least be taken into account in future research when analyzing data from a variety of perceptual learning procedures.

References

Ahissar, M. & Hochstein, S. (1993). Attentional control of early perceptual learning. Proc Natl Acad Sci USA, 90:5718-5722

Angulo, R. & Alonso, G. (2012). Human perceptual learning: the effects of pre-exposure schedules depends on task demands. *Behavioural Processes,* **91**, 244-252 De

Angulo, R. & Alonso, G. (2013). Attentional changes in human perceptual learning. *Behavioural Processes,* **98**, 61-68.

Crist R.E, Kapadia M.K, Westheimer G., & Gilbert C.D. (1997): Perceptual learning of spatial localization: specificity for orientation, position, and context. Journal of Neurophysiology, 78:2889-2894.

Dwyer, D.M., Hodder, K.I., & Honey, R.C. (2004). Perceptual learning in humans: Roles of pre-exposure schedule, feedback, and discrimination assay. *The Quarterly Journal of Experimental Psychology, **57**B* (3), 245-259.

Fahle, M. & Poggio, T. (2002). *Perceptual learning.* Cambridge, MA: MIT Press, Bradford Books.

Gibson, E. J. (1969). *Principals of perceptual learning and development.* New York: Appleton-Century-Crofts.

Gibson, J.J. & Gibson, E. J. (1955). Perceptual learning: differentiation or enrichment? *Psychological Review, **62***, 32-41

Gibson, E. J., Gibson, J. J., Pick, A. D. & Osser, H. (1962). A developmental study of the discrimination of letter-like forms. *Journal of Comparative Phisiolpgy and Psychology, **55**,* 897-906.

Gilbert C, Ito M, Kapadia M, Westheimer G: Interactions between attention, context and learning in primary visual cortex. Vision Res 2000, 40:1217-1226.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology, **49***, 585-612.

Hall, G. (2001). Perceptual learning: Association and differentiation. In: Mowrer, R.R., Klein, S.B. (Eds.), Handbook of contemporary learning theories. Erlbaum, Mahwah, NJ, pp. 367–407.

Hall, G. (2003). Learned changes in the sensitivity of stimulus representations: Associative and nonassociative mechanisms. *The Quarterly Journal of Experimental Psychology, **56**B*, 43-55.

Honey, R.C., Bateson, P., & Horn, G. (1994). The role of stimulus comparison in perceptual learning. *Quarterly Journal of Experimental Psychology, **47**B*, 83-103.

Honey, R.C. & Hall, G. (1989). Enhanced discriminability and reduced associability following flavor preexposure. *Learning and Motivation, **20***, 262-277.

Ito M., Westheimer G., & Gilbert C.D.Attention and perceptual learning modulate contextual influences on visual perception.Neuron 1998, 20:1191-1197.

Lavis, Y. & Mitchell, C. (2006). Effects of pre-exposure on stimulus discrimination: An investigation of the mechanisms responsible for human perceptual learning. *Quarterly Journal of Experimental Psychology,* **59**, 2083-2101.

Mackintosh, N.J., Kaye, H., & Bennett, C.H. (1991). Perceptual learning in flavour-aversion conditioning. *Quarterly Journal of Experimental Psychology, **43***B, 297-322.
McLaren, I.P.L., Kaye, H., & Mackintosh, N.J. (1989). An associative theory of the representation of stimuli: Applications to perceptual learning and latent inhibition. In R.G.M. Morris (Ed.), *Parallel distributed processing: Implications for psychology and neurobiology.* Oxford: Clarendon Press, 102-130.
McLaren, I. P. L. & Mackintosh. N. J. (2000). An elemental model of associative learning: I. Latent inhibition and perceptual learning. *Animal Learning and Behavior, **28**,* 211-246.
Mitchell, C. & Hall, G. (2014).Can theories of animal discrimination explain perceptual learning in humans? *Psychological Bulletin*, Vol **140**(1), 283-307
Mitchell, C., Nash, S., & Hall, G. (2008). The intermixed-blocked effect in human perceptual learning is not the consequence of trial spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, **34***, 237-242.
Mundy, M. E., Honey, R.C. & Dwyer, D.M. (2007). Simultaneous presentation of similar stimuli produces perceptual learning in human picture processing. *Journal of Experimental Psychology. Animal Behavior Processes*, **33**, 124-138.
Mundy, M.E., Honey, R.C., & Dwyer, D.M. (2009). Superior discrimination between similar stimuli after simultaneous exposure. *Quarterly Journal of Experimental Psychology*, **62**, 18-25
Nelson, J.B. (2009). Perceptions of perceptual learning. [Editorial]. *Learning & Behavior, **37**, 119.*
Pick, A. D. (1965). Improvement of visual and tactual form discrimination. *Journal of Experimental Psychology*, **69** (4), 351-339
Rescorla, R. A., & Wagner, A. R. (1972). *A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement.* In A. H. Black & W. F. Prokasy (Eds.), Classical conditioning II: Current research and theory. New York: Appleton-Century-Crofts, 64-99
Seitz, A.R. & Dinse, H.R. (2007). A common framework for perceptual learning. Current opinion in Neurobiology, 17:1-6
Seitz A.R. & Watanabe T. (2005). A unified model for perceptual learning. Trends in cognitive sciences, 9 (7):329-334
Seitz A.R. & Watanabe T. (2003). Psychophysics: Is subliminal learning really passive? Nature, 422:36
Shiu LP, Pashler H: Improvement in line orientation discrimination is retinally local but dependent on cognitive set.
Symonds, M. & Hall, G. (1995). Perceptual learning in flavour aversion conditioning. *Learning and Motivation, **26**,* 203-219.
Symonds, M. & Hall, G. (1997). Stimulus preexposure, comparison, and changes in the associability of common stimulus features. *The Quarterly Journal of Experimental Psychology*, **50**B, 317-331.
Wagner, A. R. (1981). SOP: A model of automatic memory processing in animal behavior. In N. E. Spear y R. R. Miller (Eds.). Information processing in animal: Memory mechanisms. Hillsdale, NJ: Erlbaum, 5-47.

Wang, T. & Mitchell, C. (2011). Attention and relative novelty in human perceptual learning. *Journal of Experimental Psychology: Animal Behavior Processes, **37** (4)*, 436-445

Wills, S. & Mackintosh, N. J (1999). Relational learning in pigeons? *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology*, 52(B), 31-52.

Figure Captions

Figure 1. Stimuli presented in Experiment 1. All the stimuli had 5 characters. The stimuli differed by one distinctive character, with the other four elements being common to all. In the first block of trials of Experiment 1 the stimuli were presented exactly in the order shown here.

Figure 2. Examples of the stimuli presented in Experiment 2. The short high distinctiveness stimuli had 5 characters, with only one being common to all the stimuli, and the other 4 being different in each stimulus. The long low distinctiveness stimuli were constructed from the short low distinctiveness stimuli, adding 5 common elements to these. Thus, stimuli differed in one character among 10. Finally, the long high distinctiveness stimuli also had 10 characters, only one being common to all the stimuli, and the other 9 different in each.

Figure 3. Percentage of errors (Mean ± SEM) across the three blocks of 20 trials, for the four groups of Experiment 2.

Figure 4. Percentage of same/different errors (Mean ± SEM) across the three blocks of trials, for the two groups of Experiment 3.

Figure 5. Percentage of same/different errors (Mean ± SEM) across the three blocks of trials, for the four groups of Experiment 4.

Figure 1

Figure 2

Figure 3

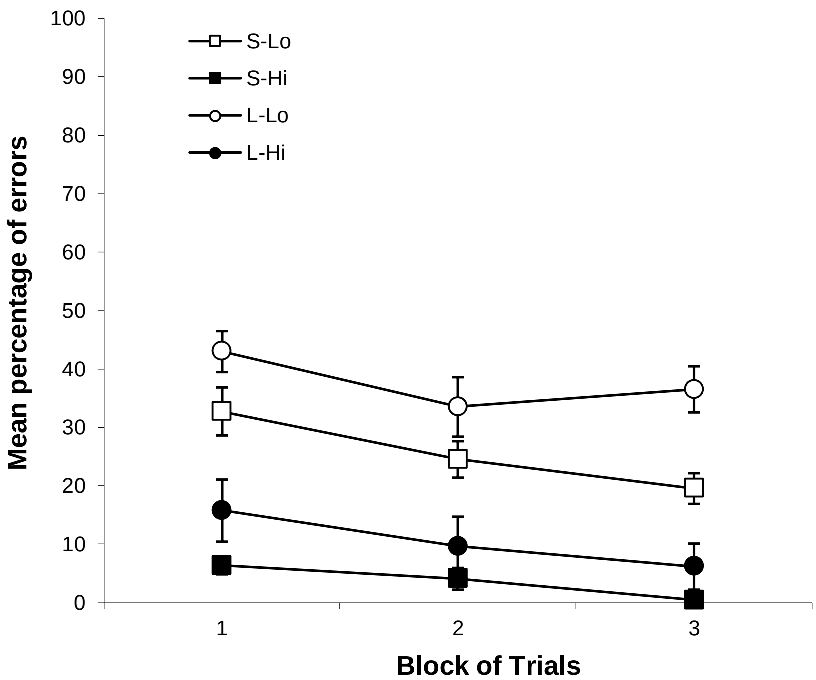

Figure 4

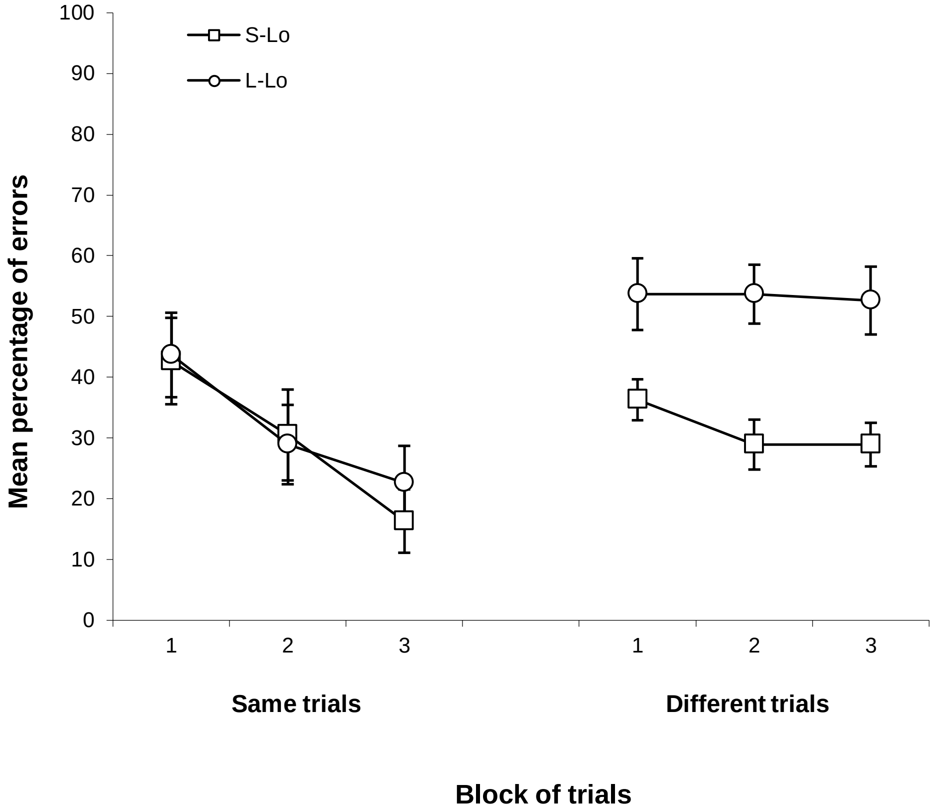

Figure 5