An Analogue of The Phillips Effect

Mark Suret (m.suret@psychol.cam.ac.uk)
Department of Experimental Psychology; Downing Street
Cambridge, CB2 3EB. UK

I.P.L. McLaren (iplm2@cus.cam.ac.uk)
Department of Experimental Psychology; Downing Street
Cambridge, CB2 3EB. UK

Abstract

Previous experimental work has demonstrated that human participants can easily detect a small change in a visual stimulus if no mask intervenes between the original stimulus and the changed version and the inter-stimulus interval is small (Phillips, 1974). Rensink, O'Regan & Clark (1997) have shown that if a mask is used then detecting the change is extremely difficult, no matter how small the ISI is made. This work attempts to establish whether familiarity with a stimulus has any effect on a participants ability to detect a small change in it using Rensink's masking procedure with Phillips' stimuli (checkerboards). Participants were required to make judgements as to whether two stimuli, which alternated with one another in presentation, were the same or different. Some participants attempted the task using just one checkerboard pattern which became increasingly familiar across sessions, others were given new, randomly generated checkerboards for each trial. In both conditions, any change (which would occur on 50% of trials) would only affect one square of the pattern. The results show a clear advantage for the participants dealing with familiar stimuli in detecting any change, and go some way towards explaining why this is so.

Introduction

Phillips (1974) demonstrated how easy it was for participants to detect a change between two stimuli if they were presented one after the other without a gap in a single alternation. This is the Phillips Effect. He also investigated the consequences of inserting a grey mask and a blank screen between the two stimuli. The inclusion of an inter-stimulus interval adversely affected participants' performance, and the presence of a mask made performance even worse. Rensink et al. (1997) demonstrated that the brief inclusion of a grey mask between repeated presentations of two slightly different stimuli made any change extremely difficult to detect. This is the Rensink Effect. Their experiment used electronically altered images which allowed manipulation of the colour, position and presence of an object. Without the mask, spotting the difference becomes trivial, if the stimuli are positioned in the same place and then alternated. Current explanations of this phenomenon cite retinal transients (Klein, Kingstone & Pontefract, 1992) as the mechanism for detecting changes in this latter case, which would be unaffected by familiarity.

Some pilot work using a single participant indicated that the effect of familiarity with the stimuli was likely to be very significant. Hence introducing the notion of familiarity removes some of the difficulties intrinsic to the Rensink Effect and makes the task more similar in difficulty to the Phillips Effect. One drawback of this pilot experiment was that it contained familiar and random trials in each session so after a while the participant became able to tell which were the familiar trials, and this may have differentially affected the responses to each trial.

Nevertheless, the results of the pilot experiment (Figure 1) allowed the prediction that the Familiar condition would lead to better detection of changes than the Random condition. It was also predicted that there would be an improvement in performance as the amount of time spent on the task increased.

The main aim of this work was to take the Rensink Effect, and attempt to ameliorate it, by allowing participants to practice on the same stimulus all the time. This would require the use of a different type of stimulus, as the repeated use of a real life scene would be impossible to control properly, so checkerboards were used as the training stimuli. In this way, the participants were presented with essentially the same stimulus on every trial, but with the possibility of a change in one of the elements within the

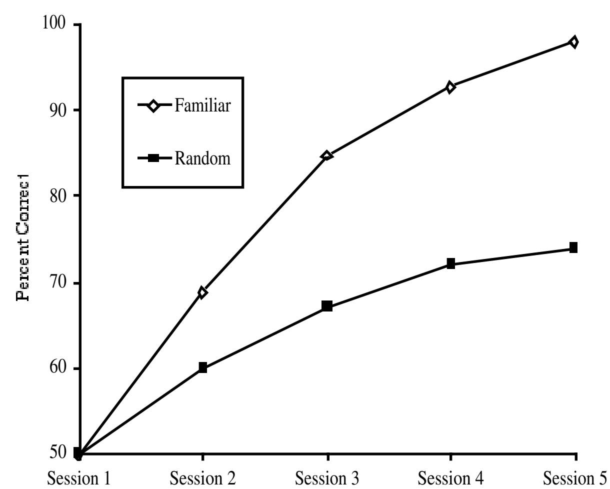

Figure 1: The Basic Familiarity Effect

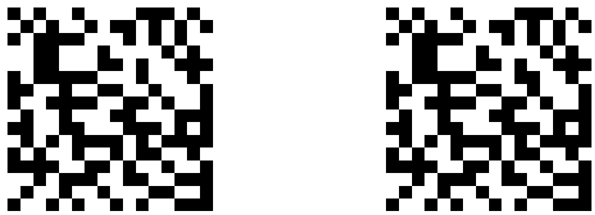

Figure 2: An example pair of chequerboard stimuli.

checkerboard. An example pair is shown in Figure 2) with the difference being rather difficult to spot.

Once a reliable difference had been found between the Familiar and Random groups, subsequent testing concentrated on the mechanism being used by participants to detect the changes.

Experiment

The experiment was conducted with two groups, each with different sets of stimuli. The Random group was a control group, and received different, randomly generated checkerboards for each trial during the experiment. In the Familiar group, each participant was trained on a single checkerboard unique to that participant. The primary aim of the experiment was to determine whether familiarity with the stimulus affected the participants' performance.

Initially two participants were run in each condition followed by a preliminary analysis. It was found that a large difference between the groups had already been established, and the final four participants were all run in the familiar condition. This allowed manipulations in the familiar condition, namely test blocks which departed from the standard task. The test blocks used manipulations intended to disrupt the performance of the participants in the Familiar group in the hope that this would suggest a possible mechanism for the way the changes were being detected. Clearly an effect of familiarity or session needed to be established first, as there is no directly relevant and properly controlled research in this area.

Stimuli and Apparatus

All the stimuli were randomly generated, two centimetre square checkerboards, with sixteen elements on a side giving a total of 256. Each base pattern stimulus had equal numbers of black and white squares, before any change was introduced. An example pair for the change condition is shown above (Figure 2), with the difference between the two checkerboards in the top right centre of the stimulus. Checkerboards were chosen as they are easy to manipulate for this type of experiment. Many different individual changes could be made whilst keeping the majority of the stimulus the same. In addition, the participants were unlikely to be familiar with the stimuli prior to the experiment

Those participants assigned to the random condition were given a newly generated checkerboard on each trial, whereas those in the familiar condition were always presented with the same pattern, albeit with a change on half the trials.

The experiment was run in a quiet room on an Apple Macintosh LCIII computer using a colour monitor. Participants responded to the stimuli by pressing one of two keys, either [x] or [.] on a QWERTY keyboard. Between blocks, participants were required to fill in a sheet to record their errors and reaction times before pressing the space bar to continue to the next block. The responses for each block were logged in separate data files.

Participants and Design

In total eight Cambridge undergraduates took part in the study. Four were allocated to the initial phase to determine the possible existence of a familiar/novel distinction. Two participants were allocated to the novel condition and two to the familiar condition. The remaining four participants were all allocated to the familiar condition.

The experiment consisted of a training phase for all participants and a test phase for those participants in the familiar group. The first four sessions were used for training for both groups, with the fifth session containing some test blocks for participants in the Familiar condition. All sessions for the random group were identical, as the tests given to the familiar group would have made no difference to a participant receiving a new checkerboard on every trial. Each one hour session consisted of ten blocks of stimuli, each containing 24 trials giving 240 trials in a session. Each block contained twelve trials where a change was present and twelve where there was no change between the two checkerboards. These trials were presented in a random order.

Participants were asked to try to detect a change between the two checkerboards, and respond appropriately as to whether or not they thought that a difference was present. For each trial, two checkerboards were alternated with one another, separated by a randomly generated mask. These checkerboards were either the same, or differed by one element within the pattern, i.e. one element that was black in one checkerboard was white in the other. The trials were such that each checkerboard was displayed for 500 milliseconds and the mask for 100 milliseconds. Over one trial, each checkerboard could be presented ten times, giving nineteen changes in a trial if the checkerboards were different. After the final alternation, the trial ended and if no decision had been made by the participant, then they were timed out.

During the fifth session, the participants in the familiar group were given test blocks in between familiar blocks, in an attempt to determine how they might be detecting the changes. These test blocks were ones containing random trials, such as those given to the participants assigned to the random group, and another type of block, labeled "C". In these blocks, there was always a fixed, random one square difference from the original base pattern, on both checkerboards, whether the trial was one of change or no change. This fixed change was different on each trial within the block. On change trials, there was also an additional change made to one of the checkerboards. This manipulation ensured that some difference from the base pattern was no longer a cue for change, although there was still a single change present between the two checkerboards on change trials. The idea behind this manipulation was to contrast any changes in performance on “C” trials with that obtained on Random trials. In the former case, the perturbation of the familiar pattern is minimal, in the latter case it is, in some sense maximal. The sequence of blocks was: Familiar, “C”, Familiar, Random, Familiar, “C”, Familiar, Random, Familiar, “C”. This gave three Familiar (as the first block is removed from any analysis), three “C” and two Random blocks to be used in the analysis for each participant from a session of ten blocks. The Familiar blocks were inserted between the test blocks to allow the participants an opportunity to re-establish baseline performance before the next test block was administered.

Procedure

The participants were seated in front of the computer approximately 50 centimetres from the screen and asked to make themselves comfortable. They were then read the instructions concerning their task, and were then asked if they had any questions about the instructions they had just been given. The participants were asked to respond to a “change” trial with their left index finger, by pressing the [x] key, and to a “no change” trial by pressing the [.] key with their right index finger. The participant was then asked to press the space bar to begin, and follow the on screen instructions that occurred throughout the experiment. The experimenter waited in the room until the first few trials had been completed, to ensure that the participant fully understood what it was that they were meant to be doing before leaving the room. Each block was started by pressing the space bar. The trials consisted of the alternation of two checkerboards, with a random black and white dot pattern mask being presented between presentations of the checkerboards. These checkerboards were either the same or differed by one element. The checkerboards subtended a visual angle of approximately two degrees and were presented in the centre of the screen. The participants were given feedback on each trial, with the words “correct” or “error” being displayed on the screen. If an error was made, the computer also beeped. After each block of twenty-four trials, participants were required to record their errors and reaction times on a sheet provided for them in the room. This was primarily to get the participants to take a break between blocks. It also gave a readily available source of data that could be tallied with the analysis on the computer. At the end of the session of ten blocks, the participants were given a short questionnaire to determine how motivated they were feeling during the session and what, if any strategy they were using.

After the questionnaire had been completed the participants were thanked for their time and the next session was arranged. After the fifth session, a more thorough questionnaire was administered, and the participants were paid and thanked for their participation.

Results

The basic familiarity effect is shown below in Figure 3. The Familiar and Random groups are denoted by F and R respectively. The graph shows that both groups improved at the task at roughly the same rate, but that performance in the Familiar group is better than that in the random group on all sessions by a roughly constant amount.

The initial analysis focused on finding significant effects of both group and session. Three dependent variables have been determined for each session: overall accuracy; percentage of changes detected; percentage of correct no change trials. Each of the three variables used may indicate something different about the way that the participants may be performing their task.

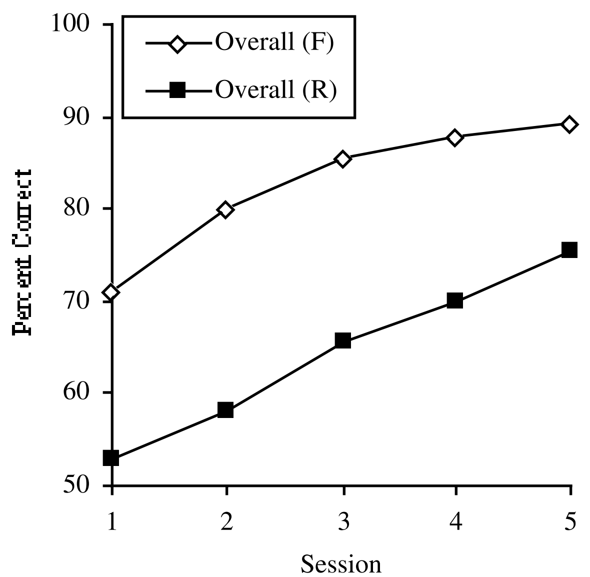

Figure 3: Between Groups Familiarity Effect.

Table 1: Within Session Comparisons between Random and Familiar Groups

(All probabilities are one-tailed)

	Overall Percentage Correct	Percentage of No Change Trials Correct	Percentage of Changes Detected
Session 1	U=0, p=0.022	U=0, p=0.023	U=0.5, p=0.032
Session 2	U=0, p=0.022	U=0, p=0.020	U=1, p=0.046
Session 3	U=0, p=0.022	U=0, p=0.022	U=0, p=0.022
Session 4	U=0, p=0.022	U=0, p=0.020	U=0, p=0.023
Session 5	U=0.5, p=0.033	U=0, p=0.017	U=1, p=0.048

Table 2: Between Session Comparisons Collapsed across Groups

(All probabilities are one-tailed)

Change between Sessions	Overall Percentage Correct	Percentage of No Change Trials Correct	Percentage of Changes Detected
1 and 2	T=1, p=0.009	T=5, p=0.034	T=2, p=0.013
2 and 3	T=0, p=0.009	T=2, p=0.036	T=0, p=0.014
3 and 4	T=2.5, p=0.024	T=6.5, p=0.100	T=6, p=0.046
4 and 5	T=6, p=0.173	T=0, p=0.008	T=8, p=0.156

Only non-parametric tests were used to analyse the data, as there were both small and unequal numbers in the groups. A significance level (α) of $p<0.05$ was used for all analyses. In each session, the score from the first block was not included in any subsequent analysis. This was to allow the participants an opportunity to practice the task before the session proper began. Means for each variable for a given session were used for all analyses. In the session 5 analyses, the different block types were separated and then the average for each variable was used.

Differences between Groups

This analysis compares the performance of the two groups on all three variables. For session five, only the familiar blocks are used for the analysis. The Mann-Whitney U-Test is used to compare the two unrelated samples.

The effect of group is so large that the difference between the Familiar and Random groups is significant for all variables in all sessions. In each case the percentage is higher for the Familiar group. A summary of the statistical results obtained from the analyses for comparisons between the Familiar and Random groups for each session is given in Table 1. Probabilities reported are one-tailed following the results of the pilot work discussed in the introduction.

Differences by Session

The test used in these analyses was the Wilcoxon Matched Pairs Test, as the values being compared were from the same participant, tracking their improvement as the sessions progressed (Table 2).

The effect of session is also a major factor in the performance of all the participants. Early in training there is the greatest effect of session, where participants rapidly become more familiar with their task and are able to improve easily. In later sessions, ceiling performance is being approached, so the difference between subsequent sessions will not be as great and the effect will be reduced.

Results from Session 5

The differences between the variables for the different blocks were compared. The comparisons of greatest interest were between the familiar and "C" test blocks and the performance of the participants in the familiar group on the random blocks compared with the performance of the participants trained solely on random stimuli. The predictions made for the effect of the test are that the participants' performance will be worse on the test blocks than on the familiar blocks presented during the session. This prediction was made as the familiar stimulus which the participants have been trained was to be distorted, to a greater (Random blocks) or lesser ("C" blocks) extent in session 5. It was deemed extremely unlikely that these manipulations would improve performance, hence one-tailed probabilities were used.

Figure 4 shows the data obtained in session 5. The Familiar, "C" and Random entries in the histogram are the averages for each block type collapsed across all the participants in the Familiar group. The control group has been included so that their performance can be compared with that of the Familiar Group on the Random blocks.

When comparing the familiar with the random blocks, all three variables produced significant results. The familiar blocks produced higher scores in the overall percentage correct (T=0, p=0.014) and the percentage of change (T=0, p=0.014) and no change trials correct (T=0, p=0.015). The Random group have been trained on the general task of detecting one square changes within a randomly generated checkerboard. The Familiar group

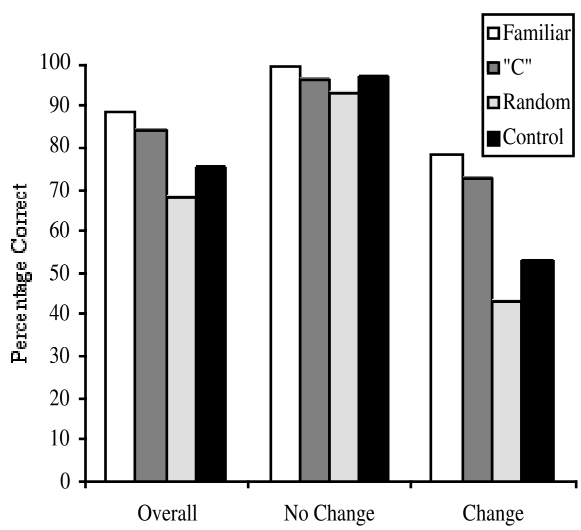

Figure 4: Comparisons between different blocks during Session 5.

have been trained on a task which can be considered as much more specialised. Using the task on which the Random group have been trained allows the comparison between the their performance and the performance attained by the Familiar participants to see if there is any overall advantage of training on a unique stimulus. No significant differences were found for any of the three variables when comparing the performance of the participants in the random group with the scores obtained by the participants in the familiar group on their test blocks with randomly generated stimuli. If anything, the familiar group participants were worse on these trials (Figure 4).

Comparing the results from the familiar and "C" trials, the only significant difference was found between the scores for the percentage of no change trials correct ($T=1$, $p=0.039$) with performance on the familiar trials being better. The other variables were found not to have any significant difference between them. Figure 5 shows the elevated false alarm rate for the "C" condition when compared with the Random and Familiar trials in session 5 and also the control group. The difference between the "C" condition and other blocks is significant in both cases ($T=0$, $p<0.05$). For this analysis, only trials where a response was made are included, thus removing any occasions on which a decision was not made in time.

Figure 5 illustrates the differences in false alarm rates for the four different cases from session 5. These rates for the Familiar participants, using their familiar stimulus, and the control group are on zero, as they registered no false alarms.

Discussion

The effects of session and group have been found to be significant. Although this is not in itself surprising, these two effects had to be established before any further investigations could be carried out. The main effect of group was surprising in its magnitude, with the effect being carried through the sessions. If the participants were allowed to reach asymptotic performance, the prediction from the pilot study is that this significant difference would be maintained. However, this prediction cannot be confirmed as the Random group were not trained for long enough to allow their performance to plateau.

Although the effect of group may have been expected with it being easier to detect changes in a stimulus that is familiar, it could have been the case that there was no difference. If the participants were simply detecting the change between the two checkerboards, then there might have been no difference in the scores. It could have been the case that the unchanging part of the pattern was irrelevant as it was the change between the checkerboards which was being probed. However, the demonstration of a difference between the groups implies that there must be another process at work, apart from the mechanism being used to detect the changes themselves. This position is supported by the finding of no significant difference between the two groups of participants when tested on the random stimuli in session five. Whatever the mechanism

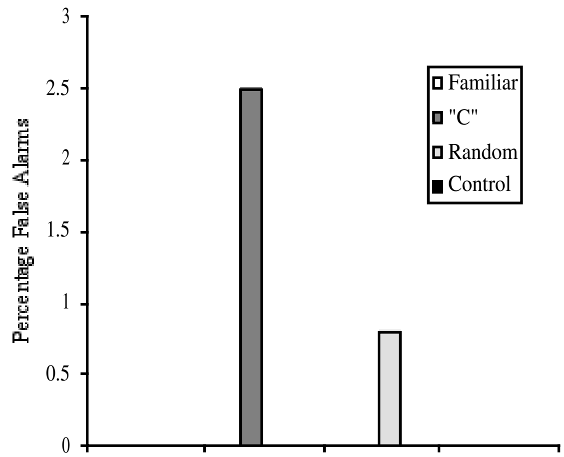

Figure 5: Comparisons of False Alarm rates during Session 5.

favouring the Familiar group, it is specific to the familiar stimuli and not some global strategic advantage developed during training.

The questionnaires administered after the sessions give an insight into how the participants were trying to detect the changes. Every participant reported using a scanning strategy, starting in a particular place for each trial, and then working around the checkerboard, attempting to spot the change. Participants in the random group employed this method throughout their sessions, with no modifications. Participants allocated to the familiar group also used such a strategy, but they reported some modification in later sessions. The scanning became faster, whilst still improving in accuracy between the sessions, there was also a chance to learn about the pattern. They were able to divide the stimulus into sub-patterns and detect changes in these, rather than searching for a single change. This certainly seems to make the detection of differences easier, as performance for the Familiar group was significantly higher for all variables during all five sessions.

The test session produced a significant difference between the percentage of no change trials detected for the "C" blocks and the familiar. In the "C" blocks, there was always a random one-square difference in the checkerboard pairs from the original, familiar stimulus. This additional fixed difference made sure that novelty from the familiar was no longer a necessary signal that there was a difference between the two checkerboards. Many of the participants reported that they had noticed the additional fixed change, and were consciously trying to avoid mistaking it for the actual change between the checkerboards. This result implies that novelty, in addition to a strategic search is central to the task being performed by the participants in the Familiar group. If it were solely a search strategy, then there should be minimal disruption of performance on test blocks which involve a minimal change in the familiar stimulus. The analysis of the false alarms (Figure 5) supports this position, as there was a greater disruption to the participants on the "C" trials than on the Random trials, where the stimulus was completely different from the one that they were familiar with, and despite the fact that overall performance was significantly lower on the Random trials.

Conclusion

There is good evidence that stimulus familiarity makes it easier to detect a change in that stimulus when compared with participants trained on random stimuli. This is in addition to the necessary attentional requirements reported in Rensink et al. (1997). The ability to detect the change improves with increasing familiarity with the stimulus. However, this ability is limited to that particular stimulus, or ones very close to it, and there is no advantage for other random stimuli of the same type as shown by the test blocks given in session 5.

The mechanism for the detection of the change may be based, at least in part, on novelty. The use of a distracting fixed change induces more false alarms on no change trials than would otherwise be expected. This result indicates that the strategy of scanning the image for change cannot completely account for the enhanced ability to detect the changes. The method for detecting the changes in a given stimulus may involve a combination of better stimulus scanning and the use of novelty to discriminate between familiar stimuli. It is likely that both mechanisms are important in the detection of change. The degree to which each is employed will doubtless depend on both the type and familiarity of the stimulus.

References

Klein, R., Kingstone, A., & Pontefract, A. (1992) Orienting of Visual Attention. In K. Rayner (Ed.) *Eye movements and visual cognition: Scene perception and reading* (pp. 45-65). New York: Springer.

Phillips, W.A. (1974) On the distinction between sensory storage and short-term visual memory. *Perception and Psychophysics, 16,* 283-290

Rensink, R.A., O'Regan, J.K. & Clark J.J. (1997) To See Or Not To See: The Need for Attention to Perceive Changes in Scenes. *Psychological Science, 8,* 368-373