The Habituation Effect in Personnel Security Polygraph Screening

Terence J. Thompson and Amy B.¹

The habituation of polygraph responses has been an object of prior research by several scientists. Most of the extant polygraph research, however, has been completed in the domain of something other than screening polygraph. Indeed, this has been the trend in polygraph research dating back to the inception of polygraph as a tool for detecting truth and deception. In today’s world, however, the vast majority of polygraph tests administered for the federal government are of the personnel security screening variety. The purpose of this study is to examine the phenomenon of habituation in personnel security screening testing.

There appears to be a common perception among polygraph examiners that as the number of questions on a chart progresses, the intensity of responses is diminished. This perception is based largely on an abundance of accumulated experience but the research record (in specific issue testing as well as other modalities) is mixed. Indeed it is reasonable to speculate that habituation might occur more frequently and in greater degree in Relevant/Irrelevant (RI) testing as the absence of comparison questions could lead to a more relaxed psychological state of the examinee—at least in the case of the truthful subject.

The issue has significance in that none of the scoring models used in screening testing mandate placement of relevant questions in a specific order. While it is true that the placement of comparison questions, irrelevant questions, sacrifice relevant questions, etc., all have an order stipulated, the sequence of relevant questions is left to the discretion of the particular examiner. Ergo, according to the notion of habituation, were a relevant question consistently asked toward the end of a set of questions, the reactivity of the question will be diminished by virtue of its placement at the end of the set. This diminution will be unrelated to the truth or deception condition. Of course the inverse might apply as well, i.e., as a question is asked closer to the beginning of the string, the more reactive it might be. The outcome of this possible failing is a skewing of polygraph results. At a minimum, examiners ought to be aware of the consistency and intensity of this phenomenon—if it exists at all in screening polygraph.

As previously indicated the phenomenon of habituation has produced mixed research findings. Most studies that examined habituation or related response capacity (and used polygraph based scenarios) did not employ screening formats.

One exception to the above is Grimsley and Yankee (1994) who examined the effects of repeated PDD tests using the RI format on accuracy. They found a non-significant difference between tests one and two but no difference between tests one and three. Their conclusion was that overall accuracy is enhanced by administration of multiple examinations. This suggests that habituation was not a significant factor in their study.

Nakayama and Kizaki (1990) employed a Guilty Knowledge Technique (Lykken, 1959) and found that the skin conductance response to both critical and non-critical items tended to diminish. Conversely, the intensity of response when the critical item was randomly inserted into a set of questions tended to increase. They concluded that there was sufficient stimulus value to discriminate critical items from non-critical items.

¹ To contact these authors, direct inquiries to the APA Editor at Editor@polygraph.org.
Dollins, Cestaro and Petit (1998) used the Peak of Tension (POT) format in examining the effect of repeated testing on reaction levels and ultimately on test accuracy. This study, however, used a six-day interval between polygraph tests and found that, except for EDA latency responses, differential reactivity did not change significantly in repeated tests. This sustainment of reactivity was observed even in cases separated by a period of six days or more. Somewhat surprisingly, they found that EDA responses to truthful and deceptive subjects did not differ significantly—which is counter to much of the research on EDA.

Timm (1984) specifically examined EDA habituation using a mock murder scenario and a Peak of Tension format. He found that “…much of the discrepancy in the findings of laboratory detection of deception studies is attributable to differences in electrodermal habituation rates and to the factors that affect them.” Interestingly, he also found a greater degree of habituation in female subjects as compared to males.

Ben-Shakhar and Gati (2003) used both verbal stimulus and pictorial in an attempt to measure habituation. They found that in both modalities skin conductance tended to diminish. This research was performed using cases actually administered in the national security milieu. As such, ground truth was not known—except in the cases of confirmed deception and iterations of the Acquaintance Test. All cases employed the RI format and 117 charts encompassing 1008 questions were analyzed. Questions included relevant, irrelevant, key questions from the Acquaintance Test and repeated relevant questions asked on the same chart. Analysis was accomplished by dividing charts into thirds and then comparing relative reactivity of the EDA in the first third, the second third and the third third. Cases were initially selected randomly; hard copy files were blindly selected off of file shelves and examined for charts containing six, nine or twelve question sets.

Charts that had six, nine or twelve questions in the string were subsequently used by the researchers; charts that had question totals that were not six, nine or twelve were specifically deselected.

Confirmed deception questions were not selected randomly and not included in the above aggregate analysis; examiners were queried as to cases that contained known deceptions and the researchers then selected based on reviews of the overall case facts. This was done to make certain that the ostensible confirmed deceptions were categorized appropriately. Analysis of confirmed deception questions was accomplished as a separate and independent component of analysis—unrelated to the aggregate data quoted throughout this study.

EDA reactivity was measured by observation of the relative deflection of the tracing. This was calculated by noting the number of chart divisions that were navigated in response to a question. For example, if the tracing deflected three chart divisions, the question was given a value of three. The minimum fraction of a chart division calculated was .25.

Aggregate analysis shows significant habituation. When all questions were examined the aggregate diminution of the EDA signal between the first third and the second third was 76.25 and from the first third to the third third 190.25. Diminution from the first third to the last third represented a 25% reduction in signal value.

In the case of twelve and nine question strings, habituation tended to occur in about the proportion elucidated above but for six question sets the effect was slightly lower. For six question sets, 40 charts containing 240 questions were examined. Total signal value of the first third was 216.75, for the second third 205.75 and for the third third 178. There was only an 11 point drop from the first third to the second third, a 23.75 reduction from the second to the third and a 38.75 reduction from the first third to the third third. Diminution from the first third to the last third represented a 18% reduction. The average time span of the six question charts reviewed was 3.31 minutes.

For nine question sets 52 charts containing 468 questions were examined. Total signal value of the first third was 303, the second third was 278.75 and the third third was 215.25. Reduction from the first to the second segment was 24.25, from the
The Habituation Effect

second to the third was 43. Reduction from the first third to the third third was 87.75 for a percentage drop of 29%. Average time span of nine question charts was 4.25 minutes.

For twelve question sets, 25 charts containing 300 questions were reviewed. Total signal value of the first third was 227.75, the second third was 186.75 and the third segment was 164. Reduction from the first to the second segment was 41, from the second to the third segment was 22.75. The drop from the first third to the third third was 63.75 for a percentage reduction of 28%. The average time span for twelve question charts was 5.37 minutes.

In RI testing certain relevant questions are usually repeated on a chart. The researchers examined 147 pairs of repeated relevant questions. Total signal value of the first iteration was 275.50 while the total for the second iteration was 200.50. This represented a diminution of 75.50—a 27% reduction in value. Most of the first iterations occurred in the first third while the second iteration almost always occurred in the final segment of the chart. It is important to add, however, that in 11 instances EDA actually increased in the second asking of the question.

The researchers also examined irrelevant questions for the effects of habituation. We looked at 88 pairs of irrelevant questions that were repeated on individual charts (a common polygraph practice). Total signal value for the first iteration was 183 and for the second iteration the value was 120.25 for a percentage decrease of 34%.

We also examined 32 pairs of directed lies on Acquaintance tests. Total signal value for the first iteration was 96.25 while the second asking totaled 94.25. Total signal reduction was less than 1%.

Lastly the researchers looked at 27 cases which presented iterations of confirmed deception questions. In 16 cases, EDA diminished on the second iteration of the question while in 11 cases EDA actually rose on the second asking of the question. In cases of decreased signal strength, the first iterations totaled 92 while the second totaled 79.25. Total signal reduction was 14%. In the case of increased reactivity, first iteration signal strength was 21.25 while the second was 36.50 for a percentage increase of 42%.

All studies have limitations and this one was no exception. This effort was conducted using past cases administered at a national security agency. As such, ground truth was unknown with the exceptions of the confirmed deceptions, irrelevant questions and directed lies obtained in the course of Acquaintance Tests.

This endeavor did not differentiate between the impacts of time run during the chart versus number of iterations. That is, while a greater degree of habituation occurred on nine question charts versus six question charts, this trend did not persist into twelve question charts. (Habituation certainly occurred on twelve question charts but not to the degree it did so on six and nine question sets.) Moreover, it was difficult for us to determine whether habituation occurs due to time lapsed on a chart, iterations of sets of different questions or the repetition of the same question twice. Rather, it was the purpose of this study to determine IF habituation occurs in the screening mode. We will speculate why this is the case later in this writing.

It is important to note that the above data do not suggest that habituation is a universal nor near universal phenomenon. We observed many instances where the phenomenon did not occur. For example, in the case of 147 repeated relevant questions EDA diminished in 75 iterations, increased in 52 instances and remained constant in 20 instances. Clearly, examiners must determine by observation early on in the test if an individual polygraph subject is prone to habituate and act accordingly.

All observations made in this study were made within the context of one chart. Neither chart to chart nor session to session reactivity was measured.

The data in this study tend to suggest that habituation is a real phenomenon in personnel security screening testing. This finding extends most importantly to relevant
questions and repeated relevant questions but it also applies to irrelevant questions. Interestingly, it does not appear to apply to directed lies obtained in the course of an Acquaintance Test. All of these findings tend to be consonant with generalized beliefs held based on the experience of many examiners engaged in day-to-day testing.

In the case of repeated relevant questions the researchers knew that the truth or deception condition was consistent. Unlike in a string of different questions which may represent different (and unknown) truth or deception conditions, repeated relevant questions offer the same condition in each iteration. In a set of different relevant questions a subject may go from truth to deception and back again as he/she negotiates the question string. It might be expected that this variation would likely reduce habituation as the subject is drawn into a higher state of tension as he/she transgresses from the truth to deception condition. In the cases of repeated relevant questions we found the overall diminishment of the signal was about 27%. This strongly suggests that signal reduction from the first to second asking is due to habituation.

It is interesting to note that in the cases of directed lies in the Acquaintance Test habituation seems to occur hardly at all. What might account for the absence of habituation in a series of questions which lack a sense of jeopardy—like irrelevant questions? Unlike irrelevant questions, which are known truths referencing non-emotion evoking issues, the Acquaintance Test employs a confirmed deception. In addition, despite the fact that we generically advise the examinee that any question can and will be repeated, the key question likely has a surprise impact as it is the first repeated question. This is supported, in part, by the EDA increase in 11 instances of repeated confirmed deception questions.

It is also interesting to note that in the cases of directed lies the Acquaintance Test employs a confirmed deception. In addition, despite the fact that we generically advise the examinee that any question can and will be repeated, the key question likely has a surprise impact as it is the first repeated question. This is supported, in part, by the EDA increase in 11 instances of repeated confirmed deception questions.

It is also important to note the frequency of habituation. As previously stated habituation is far from a universal phenomenon. In combined analysis of repeated questions, (irrelevant questions, key questions in the Acquaintance Test and repeated relevant questions) EDA reactivity diminished in the second iteration of the question in 167 instances while it increased in 100 instances. In 27 instances there was no change in EDA reactivity in the second asking of the question.

This study appears to be consistent with the intuitive holdings of most examiners, i.e., habituation occurs in personnel security screening polygraphs but is not universal. Some subjects habituate and others do not--or do so much less. The frequency and intensity of habituation has implications for scoring charts. As a question is asked in more forward position it will tend to be more reactive; as it is asked in a later position it will tend to be less reactive. Examiners need to make early observations as to an individual’s propensity to habituate or not—and proceed with that knowledge in mind.

This study does not “prove” the notion of habituation in RI screening. The data collected herein suggests the phenomenon exists and may/should stimulate additional research to confirm the findings.
References


