Improving Vigilance Analysis Methodology: Questioning the Successive Versus Simultaneous Distinction

Article *in* Proceedings of the Human Factors and Ergonomics Society Annual Meeting · September 2015

DOI: 10.1177/1541931215591059

4 authors, including:

Daniel Gartenberg
Pennsylvania State University

Bella Veksler
Wright-Patterson Air Force Base

Improving Vigilance Analysis Methodology: Questioning the Successive Versus Simultaneous Distinction

Daniel Gartenberg[1], Glenn Gunzelmann[2], Bella Z. Veksler[2], J. Gregory Trafton[3]

George Mason University[1] Fairfax, VA

Air Force Research Laboratory[2] Dayton, OH

Naval Research Laboratory[3] Washington, DC

We describe a vigilance experiment of a successive task and a simultaneous task. Successive tasks require comparing the current stimulus on the screen to a representation in memory (i.e. making a declarative memory retrieval), whereas simultaneous tasks require making a comparative judgment based on information that is available on the screen. When analyzing the data from this experiment using conventional methods, there was an effect of time-on-task (i.e. block), an effect of task type, and an interaction between block and task type. These findings were consistent with previously reported studies regarding the successive and simultaneous vigilance task distinction, which interpret such findings as evidence that the decrement is more severe for successive tasks. But different results and conclusions are made when more appropriate analyses of the data are used, such as: including block as an interval variable instead of a categorical variable and making the dependent variable detection of critical signals instead of using A'. When these analysis techniques were used, there was no effect of task type and there was no interaction with time on task. This raises questions about many of the findings in the literature, especially those regarding the successive and simultaneous distinction.

INTRODUCTION

There is a well-known distinction in the literature between successive vigilance tasks and simultaneous vigilance tasks (e.g., Parasuraman & Davies, 1977; Davies & Parasuraman, 1982). Successive tasks require judgments relative to a target that must be stored in memory. In contrast, simultaneous tasks involve comparative judgments of elements of a single stimulus array, where all the information needed to distinguish the signal is present on the screen.

A number of studies have reported steeper decrements for successive vigilance tasks (Davies & Parasuraman, 1982, Warm & Demember, 1998; See, Howe, Warm, & Dember, 1995). In one meta-analysis, it was found that within category correlations between successive and simultaneous tasks were substantial (ranging between .60 and .80), but across category correlations were weak (ranging from .10 to .40) (Davies & Parasuraman, 1982). In yet another meta-analysis of 42 papers on the successive and simultaneous vigilance task distinction, See et al. (1995) confirmed the correlation of the vigilance decrement within successive and simultaneous categories, but not between these categories.

In these meta-analyses, the d' and A' statistics derived from signal detection theory (Green & Swets, 1966; Macmillan & Creelman, 1991) were used as the dependent variable of comparison (Davies & Parasuraman, 1982; See et al., 1995). The rationale for using d' and A' was that these metrics provided a single measure of task performance and an estimate of perceptual ability that is unaffected by the willingness to respond (Davies & Parasuraman, 1982; Warm & Jerison, 1984, See et al., 1995).

An issue with using a signal detection statistic is that doing so assumes that hits and correct rejections are represented by the same mechanism. It also makes unclear whether the effects in vigilance studies are caused by hits or by correct rejections. If different processes impact hits and correct rejections, this can result in difficulty interpreting condition differences, such as the simultaneous and successive task distinction. For example, if there is a task type effect for correct rejections, but not a task type effect for hits, the A' statistic may show a significant condition effect. However, interpreting these results as an effect of task type may not be appropriate because there was no condition effect for hits and the vigilance decrement is typically measured based on hits.

In addition to possible limitations of A' as a single measure of performance in vigilance tasks, there is another important issue with previous meta-analyses of the simultaneous and successive task distinction. In these studies, the vigilance decrement is typically measured by separating the task trials into blocks of four and then analyzing these blocks as a categorical variable in a repeated measures ANOVA. However, a "categorical variable is any variable made up of categories of objects/entities" (Field, 2009), such as baseball teams or ethnicities. Block is not a categorical variable because block is not a category, but a representation of time. This makes block an interval variable - "data measured on a scale along the whole of which intervals are equal" (Field, 2009). Therefore, it is more appropriate to analyze block an interval variable (i.e. a covariate).

From a statistical perspective, coding block as a categorical variable versus as an interval variable impacts the degrees of freedom for the ANOVA models that are typically used to analyze these data. A vigilance task with four blocks will have a degree of freedom of three if it is analyzed as a categorical variable and a degree of freedom of one if it is analyzed as an interval variable.

Additionally, analyzing block as a categorical variable can inflate the decrement because it compares each block instead of the overall trend of the blocks. For example, when block is coded as a categorical variable, when running an omnibus ANOVA on four blocks, improved performance in one of the blocks can contribute to statistical differences in the ANOVA for the block effect. This should not be the case because the vigilance decrement posits only declines in performance across block.

The current study provides an opportunity for further examination of the successive and simultaneous task distinction, while taking into account these concerns. Differences between the successive and simultaneous task condition for signal detection metrics, but not critical trials will demonstrate how correct rejections can result in misinterpretations of condition differences. If differences in how the data are statistically analyzed based on including block as a factor variable or interval variable result in different findings, this will raise important questions regarding previously reported findings related to the successive and simultaneous vigilance task distinction.

METHOD

Participants

98 University of Dayton undergraduate students participated for $15. Participation was voluntary, all participants provided informed consent, and all had normal or corrected-to-normal vision. The experiment involved collection of Transcranial Doppler (TCD) data on cerebral blood flow. TCD data are not reported here.

There were 56 males and 42 females who participated in the study. The average age of participants was 21.28 years old with a standard deviation of 3.63 years.

Data for 32 participants were eliminated because they did not perform at 60% accuracy for the practice session. One additional participant was eliminated because they started to feel sick during the task. In total, 31 participants were run in the successive task condition and 34 participants were run in the simultaneous task condition, for a total of 65 participants run.

Materials

The successive and simultaneous vigilance tasks were adapted from Szalma, Miller, Hitchcock, Warm, & Dember, (1999). Each task had 1200 trials that lasted for 2 seconds each. The 1200 trials were divided into 4 blocks, where each block of 300 had 12 critical trials randomly interspersed throughout the block, for a total of 48 critical trials. Stimuli in both conditions were presented for 200 ms followed by 1800 ms of dead-zone where the screen was blank. The task is illustrated in Figure 1.

In the simultaneous condition, the task display consisted of a large centering disk 1.4 cm in diameter flanked by two dots 0.3 cm in diameter arrayed along a horizontal vector, which passed through the center of the disk. Normally, the dots were positioned so that they both were either 1.5 cm or 1.0 cm away from the center disk. The participant was instructed to only respond when the large dot was not equidistant from the two smaller dots (Figure 1a).

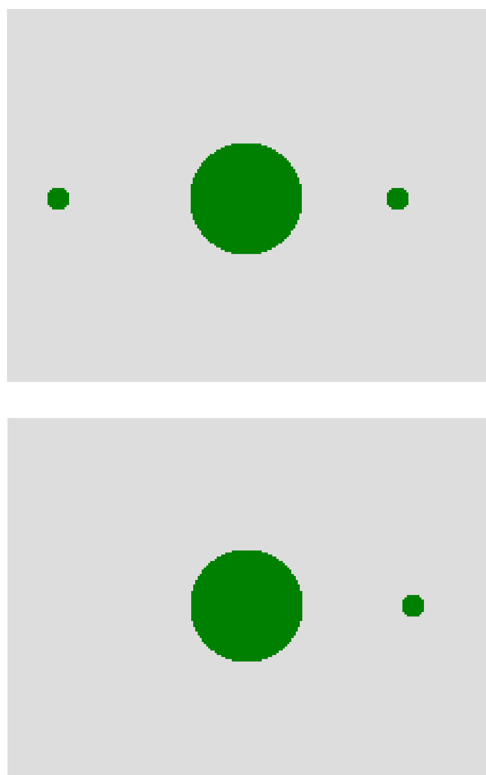
In the successive task, the display was identical to that employed in the simultaneous task, except that only a single flanking dot was used. For each presentation, the dot appeared at random either to the left or right of the disk. During neutral signal presentations, it was positioned 0.9 cm away; critical signals were cases in which the dot was 0.3 cm farther away from the center disk than usual (Figure 1b).

Design and Procedure

This experiment was a mixed factor design where the between-subjects manipulated factor was task type (simultaneous / successive) and the within subjects manipulated factor was block. There were four blocks, where each block consisted of 300 trials. Participants were assigned at random to one of the two task conditions and the experimenter configured the TCD device to the participant. Demographic information was then collected on the participant. Instructions were given to the participant on how to complete the task. The experimenter was then seated in the same room as the participant, but was separated from them by a divider. A 10-minute practice session followed. In the practice session, participants received feedback on correct or incorrect responses. Participants were trained to criteria, where if they did not perform above 60% accuracy in the first practice session, a second practice session was administered. If participants still did not perform above threshold after the second session they were eliminated from the study. After the practice, the instructions were repeated and participants were instructed to complete a 40-minute vigilance task that did not include feedback on correct or incorrect answers. Participants were then debriefed on the experiment.

Measures

Keystroke data were collected for each participant.



a)

b)

Figure 1a) *Screenshot of the Simultaneous Vigilance Task.* Figure 1b) Screenshot of the *Successive Vigilance Task.*

RESULTS AND DISCUSSION

A' Analysis

As is conventional when analyzing vigilance data, the A' signal detection measure was used (Helton, Warm, Tripp, Matthews, Parasuraman, & Hancock, 2010; Davies & Parasuraman, 1982; See et al., 1995). Unlike d', A' does not assume normality and equality of variance in the distribution of noise and signal-plus-noise.

A' is then used as the dependent variable in an ANOVA and block is typically treated as a categorical variable, where the block effect is interpreted as a vigilance decrement when A' declines. Note that block is included as a categorical variable when it ought to be included as an interval variable (Field, 2009). To demonstrate the impact of including block as a categorical variable, a mixed ANOVA was run where block was a within group categorical variable and task type (sim / succ) was a between group categorical variable. There was a statistical effect of A' for the four blocks, $F(3, 189) = 21.59$, $p < .05$, $\eta^2 = .34$. In addition, there was an effect of task type, where participants performed worse in the successive task than the simultaneous task, $F(1, 63) = 7.84$, $p < .05$, $\eta^2 = .12$. Finally, there was a statistical effect of a block and task type interaction, $F(3, 189) = 2.74$, $p < .05$, $\eta^2 = .04$ (see Figure 2).

These results are consistent with previous findings regarding the distinction between successive and simultaneous tasks, where there is typically a statistical interaction between task type and block (Davies & Parasuraman, 1982; See, at al., 1995). However, when block is a categorical variable, this interaction may not be due to performance declining at a faster rate for successive tasks (i.e. a steeper vigilance decrement).

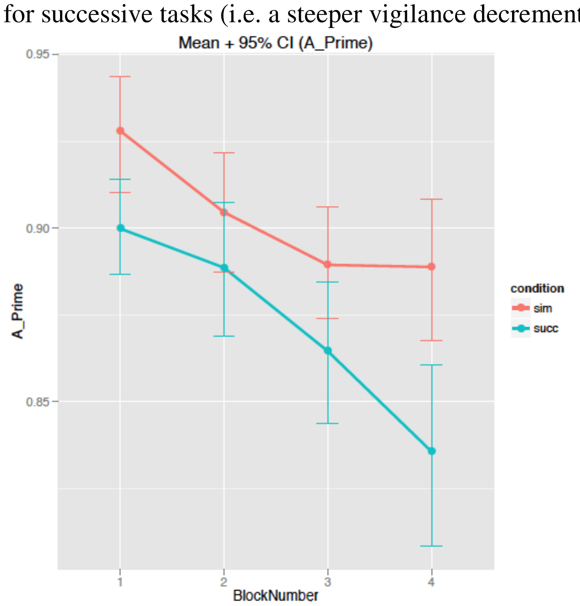


Figure 2. *A' for the successive task and the simultaneous task, over the course of the four blocks where each block represents 300 trials. Error bars are 95% confidence intervals.*

When block was analyzed as an interval variable, there was a significant effect of block, $F(1, 63) = 48.01$, $p < .05$, $\eta^2 = .85$, and again a significant task type effect, $F(1, 63) = 7.84$, $p < .05$, $\eta^2 = .12$. However, the interaction between task type and block was no longer significant when block was included as a covariate, $F(1, 63) = 2.88$, $p = .10$, $\eta^2 = .05$.

The reason that there was only a significant interaction when block was included as a categorical variable is that in these instances, each block is compared to the other block. As in this case, the effect is driven by differences in block 4 alone – instead of differences in the slope for A' between the two conditions.

When including block as a covariate, the slopes between the task type conditions are compared, meaning that overall differences in performance throughout the course of the vigil are tested. Given that the vigilance decrement is distinguished by slope differences between conditions, analyzing block as an interval variable is more appropriate. Moreover, the degrees of freedom are different when block is included as a factor variable, which further causes the different effects. Therefore, in the following analyses block will be included as an interval variable instead of a categorical variable.

Critical Trial and Neutral Trial Analysis

Often times vigilance researchers use A' to evaluate whether or not there was a vigilance decrement or condition effect. This can be misleading because A' uses both hits and correct rejections, meaning that A' effects can be driven by hits, correct rejections, or combinations thereof. When there is a task type difference for correct rejections, but no task type difference for hits, this can make interpretation of A' difficult because the vigilance decrement is typically discussed in terms of a reduction in the ability to detect critical stimuli (i.e., a reduction in hits).

This dataset is an example of when the A' measure is particularly problematic because differences in task type are driven by correct rejections instead of differences in hits (see Figure 3a and 3b). While there was a significant effect of block for hits, $F(1, 63) = 53.83$, $p < .05$, $\eta^2 = .85$, unlike the A' measure, there was not a significant effect of task type for hits, $F(1, 63) = 2.89$, $p = .09$, $\eta^2 = .05$. Also, there was no interaction between block and task type, $F(1, 63) = 1.32$, $p = .26$, $\eta^2 = .02$ (see Figure 3a).

However, for correct rejections there was not a significant effect of block, $F(1, 63) = 0.80$, $p = .37$, , $\eta^2 = .01$, but there was an effect of task type, $F(1, 63) = 9.55$, $p < .05$, , $\eta^2 = .15$. Again, there was no interaction, $F(1, 63) = 0.00$, $p = .99$, $\eta^2 = .00$ (see Figure 3b). This showed that the block effect for A' was caused by declines in percentage of correctly detected signals, but the task type effect found in A' was driven primarily by accuracy on correct rejections.

The vigilance decrement is typically indexed by declines in the percentage of correctly detected signals. If the percentage of correctly detected signals is the measure of the vigilance decrement, then differences between conditions for the decrement should also be defined by the percentage of correctly detected signals. Therefore, since there was no difference in correctly detected signals between conditions, in this example, it would be misleading to interpret the A' results as supporting differences in the vigilance decrement between the task types.

Taken together, when interpreting these vigilance tasks with the aforementioned three analyses - A', hits, and correct rejections – a vigilance decrement occurred, but there was no no interaction between task type and block. There was also no effect of task type, though with more power this marginal effect may become significant.

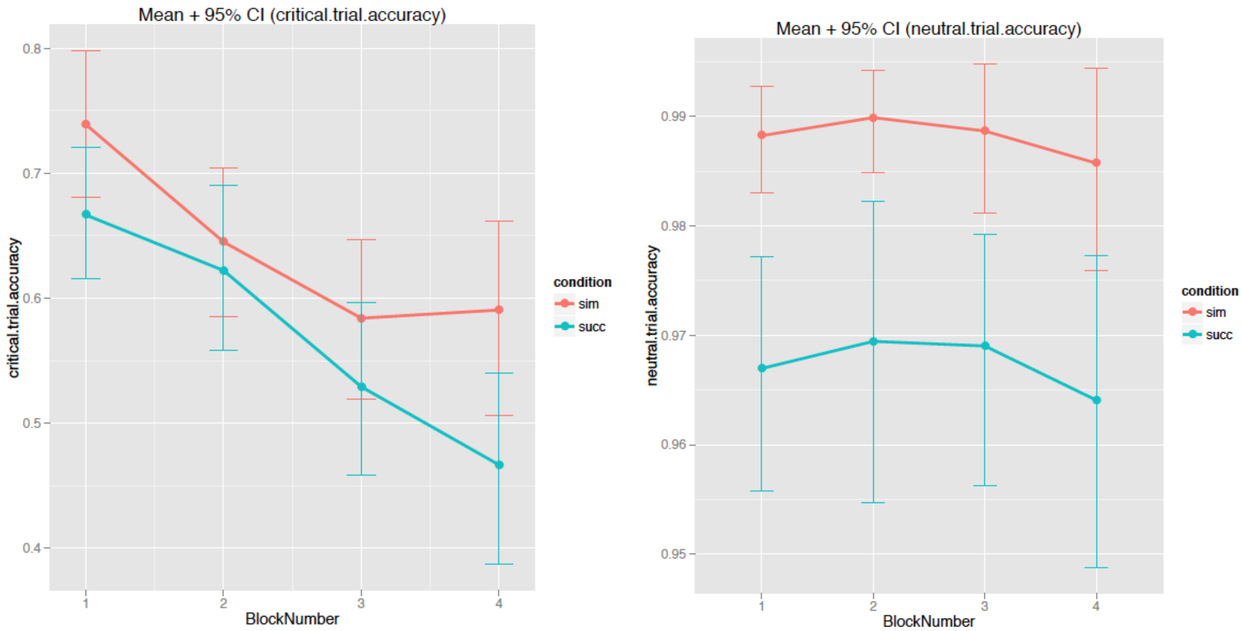


Figure 3. *3a) Accuracy on hits. 3b) Accuracy on correct rejections. Each block represents 300 trials. Error bars are 95% confidence intervals*

How data is analyzed	Result of analysis method
Block is a categorical variable	May result in statistical differences that do not truly representing a vigilance decrement. This can because if performance improves in a block, this contributes to the effect of a vigilance decrement in an omnibus ANOVA. The degrees of freedom are not incorrect.
Block is a an interval variable	Statistical differences in block represent a vigilance decrement when performance declines over the time-course of the blocks. The block variable in the ANOVA represents the overall performance trend. The degrees of freedom are correct.
A' is used as the dependent measure	Differences in condition effects can be due to either differences in hits or differences in correct rejects. If there is an effect of condition for correct rejections, but not for hits, this may not represent a difference between conditions.
Hits and correct rejections are analyzed separately	Fewer hits over the blocks represent a vigilance decrement. Differences in condition effects for correct rejections do not impact the percentage of hits.

Table 1. *A table of different ways to analyze vigilance data and how these analysis methods impact interpretation of the decrement.*

CONCLUSION

Two major methodological issues when analyzing vigilance data were identified in this paper: a statistical issue with previously conducted ANOVAs and an issue with using A' as the dependent measure (see Table 1 for a breakdown of these methodological differences and rationale)..

This paper illustrates how many of the previously reported effects in the vigilance literature are impacted by a statistical issue in how block is included in ANOVA models of task performance. Block should be included as an interval variable instead of as a categorical variable because block is a representation of time – not a category (Field, 2009).

The other issue concerns using signal detection measures, which can be problematic given that the decrement is typically defined by hits, and A' is composed of both hits and false alarms. This is particularly problematic when interpreting condition effects. Furthermore, given that that correct rejections and hits are represented by two different cognitive processes, as has been suggested by a computational model called the microlapse theory of fatigue, (Gunzelmann, Gross, Gluck, & Dinges, 2009; Gunzelmann, Moore, Gluck, Van Dongen, & Dinges, 2010; Gartenberg, Veksler, Gunzelmann, & Trafton, 2014; Veksler & Gunzelmann, *prep*), it can be problematic to interpret a single performance measure of the decrement, such as A' and d' from signal detection theory. The reason for this is that combined performance measures make the process behind what causes condition effect unclear.

When A' was used as the performance metric for distinctions in vigilance performance and block was included as a factor variable, an interaction between task type (sim / succ) and block was found. While this supported previous findings in the literature (Davies & Parasuraman, 1982; See et al., 1995), the interaction disappeared when block was included as a covariate. This is important because it suggests that the conclusions of previous studies on this topic may be influenced by the statistical analyses that were used.

Additionally, in the data reported here, the task type effect was driven by differences in false alarm performance rather than performance on critical trials. Therefore, when interpreting the condition effect found for A', it would not be correct to say that there was a difference in vigilance performance for the conditions because the decrement is typically defined by performance on hits. Therefore, the right interpretation of the data is that there was no effect of condition on vigilance performance.

Since many vigilance tasks, and in particular, studies that examine the successive and simultaneous task distinction, are evaluated based on A' and include block as a factor variable, this calls into question a defining aspect of the vigilance taxonomy: the successive and simultaneous task distinction. In order to better understand the nature of the vigilance decrement, all vigilance analyses should report both hits and correct rejections and block should be included as an interval variable in ANOVA models.

ACKNOWLEDGEMENT

This work was supported by the Air Force Research Laboratory's Human Effectiveness Directorate. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the United States Air Force or Navy.

REFERENCES

Davies, D. R. & Parasuraman, R. (1982). The Psychology of Vigilance. London: Academic Press.

Field, A. P. (2009). Discovering statistics using SPSS: and sex drugs and rock 'n' roll. London, England : SAGE.

Gartenberg, D., Veksler, B., Gunzelmann, G., & Trafton, J. G. (2014). An ACT-R process model of the signal duration phenomenon of vigilance. *In Proceedings of 58th annual meeting of the Human Factors and Ergonomics Society*

Green, D. M., & Swets, J. A. (1966). Signal detection theory and psychophysics. New York: Wiley.

Gunzelmann, G., Gross, J. B., Gluck, K. A., & Dinges, D. F. (2009). Sleep deprivation and sustained attention performance: Integrating mathematical and cognitive modeling. *Cognitive Science, 33*(5), 880-910.

Gunnzelmann, G., Moore, R. L., Gluck, K., Van Dongen, H. P. A., & Dinges, D. F. (2010). Fatigue in sustained attention: Generalizing mechanisms for time awake to time on task. In P.L Ackerman (Ed.), Cognitive Fatigue: Multidisciplinary Perspective on Current Research and Future Applications (pp. 83-96). Washington, DC: American Psychological Association

Helton, W., Warm, J., Tripp, L., Matthews, G., Parasuraman, R., & Hancock, P. A. (2010). Cerebral lateralization of vigilance: A function of task difficulty. *Neuropyschologia, 48*, 1683-1688.

Macmillan, N. A., & Creelman, C. D. (1991). Detection theory: A user's guide. Cambridge, England: Cambridge University Press.

Parasuraman, R., & Davies, D. R. (1977). A taxonomic analysis of vigilance performance. In R. R. Mackie (Ed.), Vigilance: Theory, operational performance, and physiological correlates (pp. 559-574). New York: Plenum.

See, J. E., Howe, S. R., Warm, J. S., Dember, W. N. (1995). Meta-analysis of the sensitivity decrement in vigilance, *Psychological Bulleton, 117*, 2, 230-249.

Szalma, J.L., Miller, L.C., Hitchcock, E.M., Warm, J.S., & Dember, W.N. (1999). Intraclass and interclass transfer of training for vigilance. In M.W. Scerbo and M. Mouloua (Eds.), *Automation technology and human performance: Current research and trends* (pp. 183-187). Mahwah, NJ: Erlbaum.

Veksler, B. Z., & Gunzelmann, G. (2013). Modeling the Vigilance Decrement in the Mackworth Clock Task. In D. N. Cassenti (Ed.), *Proceedings of the 22nd Annual Conference on Behavior Representation in Modeling and Simulation (BRIMS)*. San Antonio, TX: BRIMS.

Warm, J. S., & Demember, W. N. (1998). Tests of a vigilance taxonomy. In R.R. Hoffman, M.F. Sherrick, & J.S. Warm (Eds). *Viewing psychology as a whole: The integrative science of William N. Dember* (pp. 87-112). Washington, DC: American Psychological Association.

Warm, J. S., & Jerison, H. J. (1984). The psychophysics of vigilance. In J. S. Warm (Ed.), Sustained attention in human performance (pp. 15-60). Chichester, England: Wiley.