

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Counting Sheep Is a Good Way to Get to Sleep, but the Occasional Aardvark Will Wake You Up: How a Salient Event Improves Performance

Permalink

<https://escholarship.org/uc/item/9c8147pm>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 30(30)

ISSN

1069-7977

Authors

Veksler, Bella Z.

Gray, Wayne D.

Publication Date

2008

Peer reviewed

Counting Sheep is a Good Way to Get to Sleep, but the Occasional Aardvark Will Wake You Up: How a Salient Event Improves Performance

Bella Z. Veksler (zafrib@rpi.edu)

Wayne D. Gray (grayw@rpi.edu)

Department of Cognitive Science, 110 8th St.

Troy, NY 12180 USA

Abstract

Sustaining attention is a problem many of us face in various settings from the workplace to school. Sustaining attention is often crucial for good performance. *Vigilance decrement* refers to a decline in task performance as time on task increases. The current study induces a vigilance decrement across 2000 trials of a monotonous task. In this study each trial is composed of one target and one distracter presented to the participant and the participant's goal is to determine if the target is above or below the distracter. In the control condition, both stimuli are always presented in black font. For the experimental condition, on trial 1200 and 1900 the target is presented in red font. There are no other differences between conditions. This minor, task irrelevant change suffices to reduce the vigilance decrement relative to the control condition on all subsequent trials. We dub this phenomenon the *Aardvark Effect*.

Keywords: vigilance, sustained attention, alertness, performance, stimulus salience, oddball paradigm, boredom

Introduction

Sustaining attention in monotonous tasks is crucial in many areas of the workplace and in school. Because often times attention is fleeting, we would want to be able to formally quantify those instances in which attention is waning. Along the same lines, knowing which stimuli are capable of attracting and sustaining one's attention and improving performance in vigilance-type tasks is crucial in many domains.

The vigilance decrement for perceptual-motor detection and discrimination tasks is well established by over a half century of research (Bakan, 1955; Jerison & Pickett, 1963; Mackworth, 1948). It is defined as the slowing of reaction times and/or the decrease in accuracy over time on task. This decrement typically occurs about 30-45 minutes into the task (Mackworth, 1948) and is aggravated when the tasks are especially monotonous and do not require a great deal of cognitive processing (Adams, Humes, & Stenson, 1962; Adams, Stenson, & Humes, 1961). There have been several theories postulating why such a decrement occurs, citing differences between controlled and automatic processing of information, decreases in sensitivity or criterion changes as per signal detection theory, decrease in motivation, and fatigue (Fisk & Schneider, 1981; Frankmann & Adams, 1962; Jerison & Pickett, 1963; Pattyn & Soetens; Scott, 1957; Williges, 1969).

The problem of vigilance decrement can be extended to real world tasks that are not typically thought of as vigilance tasks. For example, imagine you are sitting in class listening to the lecture and after twenty-five to thirty minutes your

attention begins to wander. You begin thinking about the upcoming football game or what color to paint your living room or, worse yet, you start dozing off. Now imagine that all of a sudden, WHAM! Something unexpected happens that grabs your attention so that you become engaged in the lecture once more. This external stimulus is separable (not integral) to the semantic content of the lecture or to the current goals of task performance. Yet, somehow this occasional aardvark thrusts itself into your attention and by so doing manages to increase performance long after it has vanished.

Prior work on vigilance decrements has established the performance decline effect but has made little progress on countering it. The current study begins an effort to determine how to alleviate the waning attention that results from prolonged time on monotonous tasks. We ask whether a small and unexpected change in a stimulus occurring far into a long monotonous task could somehow increase attention and thereby reduce the vigilance decrement effect.

Background

Much of the work on vigilance began in the mid-1900's, driven, as much research is, by an applied military problem (Mackworth, 1948). Radar operators often times had to monitor radar screens for critical signals for many hours without rest. It was found that their performance in detecting critical signals worsened the more time they spent on task. Researchers were called in to study the optimal time on task that radar operators could sustain a passable level of detection performance.

Mackworth approached this problem by manipulating the duration any one operator spent on task. The task was simply to monitor a clock hand that rotated 360 degrees every minute, like a second hand typically does. The critical signal to be detected was the movement of the hand two clicks instead of the usual one per second. This double movement occurred 12 times every 20 minutes. He found that operators who did the study in half-hour spells with half hour breaks over the 2 hour period exhibited less decrement than those who remained on vigil the full two hours. This suggests, perhaps not surprisingly, that injecting breaks into a monotonous vigilance task helps maintain alertness.

Mackworth ran another condition in which participants on a 2-hour vigil of watching the clock also had to monitor a telephone message that would give them additional instructions. In fact, only once did the phone ring – at the end of the first hour of watch – and the message was to “do even better for the rest of the test.” The finding for this group was

that participants declined in detection performance over the first hour but once they heard the phone message they improved markedly. Of course it is difficult to say why this improvement occurred. Participants may have been alerted to attention by the phone ring or they may have become more motivated given the instructions of the experimenter. Regardless of the reason, it is evident that an unexpected stimulus, one clearly separable from the goals of the task, resulted in an improvement in performance.

Other researchers have manipulated the degree of the vigilance decrement by adding a concurrent secondary task, unrelated to the primary task. Not only does the secondary task increase performance in the vigilance task but it also results in the underestimation of the duration of the vigilance task and the report that it was “interesting” (Smith, Lucaccini, Groth, & Lyman, 1966). Allowing participants to engage in non-current secondary tasks such as 5 minutes of vigorous physical exercise, solving anagrams, or sensory deprivation after every half hour on the vigilance task also eliminate the vigilance decrement (Bevan, Avant, & Lankford, 1967).

Furthermore, some researchers have found gender differences during particular types of prolonged tasks (Dittmar, Warm, Dember, & Ricks, 1993; McIntosh, 2006; Prinzel & Freeman, 1997), whereas others do not (Tolin & Fisher, 1974; Waag, Halcomb, & Tyler, 1973). As an added analysis, we will look at whether there are differences in performance between males and females.

As this review shows, when intrinsic motivation wanes there are many ways in countering the vigilance decrement that are external to the task itself. Hence, adding a secondary task, providing breaks for exercise or sensory deprivation, all serve (somehow) to refresh participants so that performance on the primary task increases.

The Aardvark Effect

The current study attempts to replicate the vigilance decrement and to also show that attention can be incremented by small changes to task stimuli in cases in which these changes are irrelevant to task performance, the Aardvark Effect. Furthermore, the stimulus we use is small and the change is fleeting. The hypothesis is that the introduction of a rather moderate and goal-irrelevant change in the task will increment attention and improve performance in subsequent trials as compared to the control condition.

Method

We used a simple monotonous task lasting about 45 minutes. The manipulation was the introduction of oddball critical trials that we hypothesized would draw participants’ attention back to the task and therefore result in the maintenance of a higher level of performance than controls.

Participants

A total of 48 undergraduates (17 females and 31 males) from Rensselaer Polytechnic Institute participated in the study. The average age was 19.5 years ($SD=1.4$). Students received extra credit for their participation.

Design

There was one between-subjects independent variable producing two conditions. The variable was whether or not there were two oddball trials during the experiment. An oddball trial was defined as one in which the target letter appeared in red instead of the usual black. In all other trials, both the target and the distracter letters appeared in black against a white background (Figure 1).

Materials

The experiment was presented using a computer running Mac OS X on a 17” flat-panel LCD monitor set to 1024x768 resolution. The software used for the experiment was written in LispWorks 5.0. Each trial consisted of two letters in Arial size 100 font presented in the middle of the screen, one above the other such that there was a 2° separation between the two letters. Since participants were eye tracked during this study, the separation between letters was crucial to ensure that we could properly assign object information to fixation data. An LC Technologies eye tracker was used to collect eye data during the study at a rate of 120Hz. Participants were asked to keep their chin in a chinrest in order to ensure the eye data collected would be accurate. Current analyses do not include the eye data, however, future work will incorporate this data as a possible physiological measure of sustained attention. In particular, pupil diameter will be utilized as a measure of surprise during the oddball trial.

Procedure

Each participant was run separately. There was considerable effort given to counterbalancing the time of day participants from the two conditions were run. All participants were run in hour-long slots between the hours of 9am and 5pm. Each condition had at least one participant in each of the eight time slots (i.e., 9am, 10am . . . 4pm), in an attempt to control for any circadian rhythm effects in attention. Additionally, every effort was made to ensure that no extraneous ambient noise was present when participants were run as any outside noise could have served as a salient stimulus confounding the data we collected. For the same reason, participants were asked to turn off their cell phones for the duration of the study.

The experiment consisted of 2000 forced-response consecutive trials. On each trial, two letters appeared on the screen, one above the other and participants had to depress one of two response keys (“T” or “B”) corresponding to the location of the target letter on the screen (top or bottom). Participants were told prior to the beginning of the task that target letters were one of A, B, or C. One of these three letters was present on each trial and appeared as either the top or the bottom letter. The distracter letters were all pulled from the first 13 letters of the alphabet. There was an 800-millisecond delay between the response of one trial and the presentation of the next trial during which time there appeared a blue crosshair to sustain visual attention on the center of the screen.

After signing informed consent forms, going through the instructions on how to respond and being calibrated to be eye

tracked, each participant performed the task of responding to each of the 2000 trials. In the control condition, participants simply saw 2000 consecutive trials of target and distracter letters all presented in black font. In the experimental condition, on the 1200th and on the 1900th trials, roughly 25 and 40 minutes, respectively, into the experiment, the target letter in the trial appeared in red rather than in the usual black. Note that participants were not aware that one of the letters would change color during the study. Nothing else varied between the two conditions.

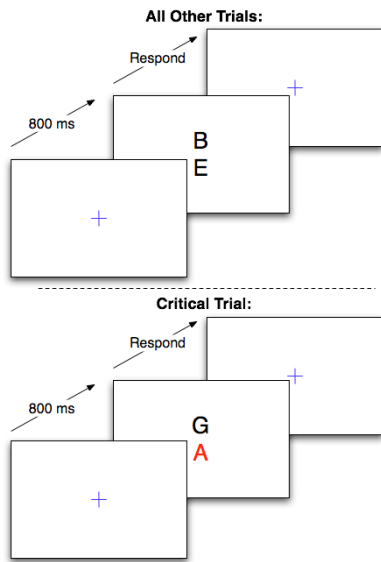


Figure 1: Experimental Stimuli

Results

Given the delicate nature of the phenomenon we are trying to study, care had to be taken to make sure no extraneous sources of stimulation occurred during the experiment. One participant was heard singing during the study, resulting in the exclusion of her data as well as that of two other participants who were run at the same time (in separate, but adjacent rooms) as they may have heard her. Several other participants had to be excluded due to a loud radiator in one experimental room and other unforeseen circumstances arising during the study. Consequently, only data from 36 participants (11 females and 25 males) was used for the analysis, 18 in each condition.

Pre- and Post-Experimental Questions

Prior to doing the experiment, participants were asked to rate their video game experience. We did this to rule out any attention-based confound associated with video game playing. However, we found that participant-reported video game experience was not correlated with overall performance on the task.

After completing the study, we asked all participants whether they had noticed anything strange during the study. All participants in the experimental condition responded that they noticed the red letter. Participants in the control

condition reported that they did not notice anything. We also asked participants, if they did notice something, whether they felt it affected their performance. Only two reported that they felt the red letter made them more alert on the critical and several subsequent trials. Other participants, especially those in the control condition, reported feeling bored and disliking the task.

Trend in Accuracy

Accuracy was measured by how many trials each participant answered correctly. That is, when the target letter was the top letter, the participant keyed in “T”, whereas when the target letter was the bottom letter, the participant keyed in “B.” We were most interested in how performance varied with respect to the critical trials.

Preliminary analyses were done with respect to the general trend of performance between the two conditions as can be seen in Figure 2. The data has been binned by 100 trials and the oddball critical trials occurred in trial bin 12 and trial bin 19 as indicated by the dotted lines on the graph.

Trial accuracy before and after the first critical trial was compared between the two conditions. To have an equal number of trials, 800 trials prior to the first critical trial were compared to the 800 trials after the first critical trial. A two-factor independent measures ANOVA on the effects of Condition and Trial Type (before or after the first critical trial) revealed a significant main effect of Condition, $F(1, 3196) = 109.65, p < 0.01$, a significant main effect of Trial Type, $F(1, 3196) = 173.24, p < 0.01$, and a significant Condition x Trial Type interaction, $F(1, 3196) = 22.00, p < 0.01$.

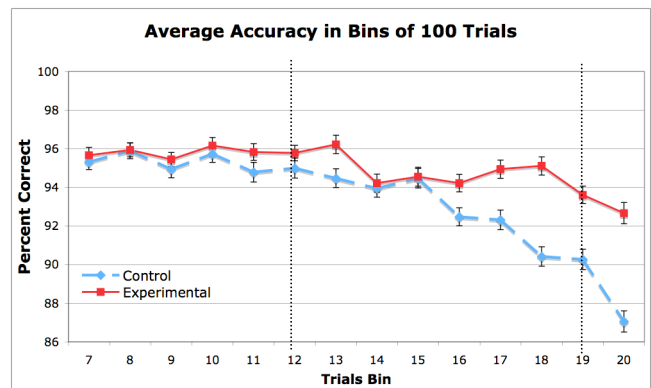


Figure 2: Trend in Accuracy Binned by 100 Trials. Error bars signify standard error.

Having done this meta-analysis, we compared accuracy with respect to the individual critical trials. Since there was considerable variability in performance throughout the 2000 trials and since the most number of trials after a critical trial was 100 (the second critical trial was the 1900th), for the present analysis we found it crucial to look at a window of 100 trials around the critical trials. In particular we ran an analysis on the 100 trials before each critical trial versus 100 trials immediately following each critical trial (Figure 3).

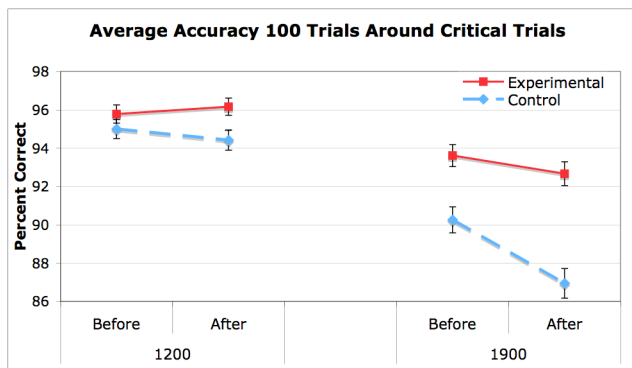


Figure 3: Accuracy Comparison Between Conditions Around the Critical Trials. Error bars signify standard error.

Trial 1200. A two-factor independent measures ANOVA on the effects of Condition and Trial Type (100 trials before or after the first critical trial, trial 1200) revealed a significant main effect of Condition, $F(1, 396) = 9.84, p < 0.01$, no effect of Trial Type, $F(1, 396) = 0.06, p = 0.81$, and no interaction, $F(1, 396) = 1.18, p = 0.28$.

Planned comparisons showed that there was no significant difference on the 100 trials before the first critical trial between the control condition ($M = 94.83, SE = 0.45$) and the experimental condition ($M = 95.78, SE = 0.44$), $t(198)_{two-tail} = -1.50, p = 0.14$. However, there was a significant difference 100 trials after the critical trial between the control condition ($M = 94.22, SE = 0.5$) and the experimental condition ($M = 96.17, SE = 0.45$), $t(198)_{two-tail} = -2.89, p < 0.01$.

Trial 1900. A two-factor independent measures ANOVA on the effects of Condition and Trial Type (100 trials before or after the second critical trial, trial 1900) revealed a significant main effect of Condition, $F(1,396) = 63.18, p < 0.01$, a significant main effect of Trial Type, $F(1, 396) = 11.77, p < 0.01$, and the Condition x Trial Type interaction was approaching significance, $F(1, 396) = 3.82, p = 0.05$.

A planned comparison of the 100 trials before the second critical trial showed a significant difference between the control condition ($M = 89.78, SE = 0.71$) and the experimental condition ($M = 93.61, SE = 0.56$), $t(187.93)_{two-tail} = -4.24, p < 0.01$. This difference had been developing as a result of the first critical trial. In addition, there was a significant difference 100 trials after the second critical trial between the control condition ($M = 86.33, SE = 0.72$) and the experimental condition ($M = 92.67, SE = 0.55$), $t(185.87)_{two-tail} = -7, p < 0.01$.

Gender Differences

Due to the mixed results of previous research on gender differences in prolonged tasks, we further analyzed the data with respect to gender. Figure 4 depicts the overall accuracy across the entire 2000 trials as a function of condition. These analyses are preliminary as we had an unequal number of males and females participating and future studies will remove this confound (Control: 6 females and 12 males; Experimental: 5 females and 13 males).

A two-factor independent measures ANOVA on the effects of Condition and Gender revealed a significant main effect of Condition, $F(1, 7996) = 61.10, p < 0.01$, a significant main effect of Gender, $F(1, 7996) = 119.85, p < 0.01$, and a significant Condition x Gender interaction, $F(1, 7996) = 70.25, p < 0.01$.

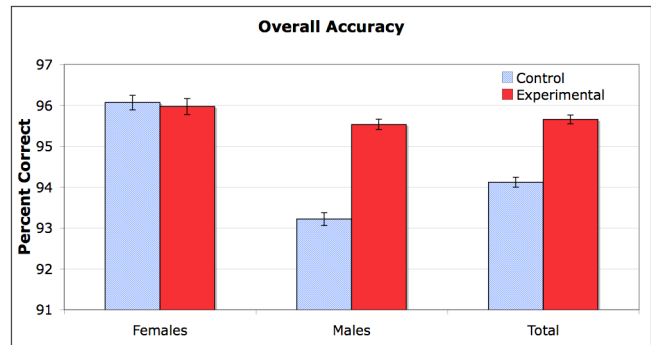


Figure 4: Overall Accuracy Across all 2000 Trials By Gender. Error bars signify standard error.

There was no significant difference for females between the control ($M = 96.07, SE = 0.17$) and experimental conditions ($M = 95.97, SE = 0.20$), $t(3918.87)_{two-tail} = 0.36, p = 0.72$. However, there was a significant difference for males between the control ($M = 92.76, SE = 0.17$) and experimental conditions ($M = 95.53, SE = 0.13$), $t(3727.51)_{two-tail} = -12.82, p < 0.01$.

A comparison of the control condition showed a significant difference between females ($M = 96.07, SE = 0.17$) and males ($M = 92.76, SE = 0.17$), $t(3998)_{two-tail} = 13.52, p < 0.01$. A comparison of the experimental condition showed no significant difference between females ($M = 95.97, SE = 0.20$) and males ($M = 95.53, SE = 0.13$), $t(3438.65)_{two-tail} = 1.84, p = 0.07$.

Discussion

We hypothesized that the insertion of a conspicuous, but goal-irrelevant, stimulus during our monotonous task would increase performance by alleviating the decrement of sustained attention. The study we ran presented participants with a small stimulus change that was completely irrelevant to task performance – the color of one target on trials 1200 and 1900. We observed that although there still was a decrement in accuracy as time on task increased, the decrement was markedly smaller for the color-change group.

Performance after the first critical trial (trial 1200), showed a large decline for the control group whereas, only a small decline was observed for the experimental group. The difference between the two conditions 100 trials after the first critical trial is about 1.75%. What is perhaps more impressive is that the decline in accuracy, the slope with which accuracy falls as a function of time, becomes different after the first critical trial (Figure 2). The slope of the accuracy curve is considerably steeper for the control group as compared to the experimental group.

Performance after the second critical trial again showed a large decline for the control group with no significant decline for the experimental group. The difference between the two groups by the end of the study is about 5.72%. This suggests that our manipulation of introducing the salient stimulus did indeed alleviate the performance decrement and resulted in 5.72% more accurate responses at the conclusion of the vigilance task.

A further analysis with respect to gender suggested that males benefited considerably more from the introduction of the conspicuous stimulus, whereas females maintained the same level of performance regardless of condition. However, this unexpected finding needs to be replicated in a study that includes equal numbers of participants of both genders. For example, others have found that depending on the type of task – spatial or temporal, there is a difference in decrement between the two genders (Dittmar et al., 1993). Taken together with prior research, the preliminary findings from our study suggest that when it comes to alleviating the vigilance decrement we may need to consider several factors with respect to who is performing the task, what time of day, what type of task it is, etc.

Conclusions and Future Directions

The current work represents an attempt to decrease performance decrements often observed in prolonged, fairly easy tasks. We achieved this effect by introducing a small change in one of the stimuli presented to participants – irrelevant to the actual task.

Our results showed that drastic measures such as sounding a loud truck horn or making participants do a secondary problem-solving task are not required to counter the performance decrement. This fairly simple and meaningless change, simply displaying one letter in red as opposed to the usual black, sufficed to maintain performance 5.72% above controls.

The key here is that the red letter served as a conspicuous and salient cue that drew participants' attention and alertness to the task. Regardless of whether they themselves felt they improved their performance after seeing the red letter, their decline was considerably shallower than controls. The aardvark (red letter) among sheep (black letters) alerted participants enough to sustain performance.

Future studies will address whether the intriguing gender differences we found are indeed real as this may indicate that varied methods may need to be employed to counter the vigilance decrement across the genders. Furthermore, the current task was used because it was the simplest and most mindless that we could think of to have our participants do. Further studies would have to investigate other types of tasks, perhaps using different modalities.

One final and crucial extension of the current work will be to operationally define the degree of conspicuity or salience required of the stimulus to sufficiently alleviate the decrement. The applicability of this work would be tremendous in many domains – especially educational and many monotonous work settings where the maintenance or

alertness is necessary for good performance but is difficult to sustain.

Acknowledgments

This work was supported, in part, by grants from the Office of Naval Research (N000140710033) and the Air Force Office of Scientific Research (FA9550-06-1-0074) to Wayne Gray.

References

- Adams, J. A., Humes, J. M., & Stenson, H. H. (1962). Monitoring of complex visual displays: III. Effects of repeated sessions on human vigilance. *Human Factors*, 4(3), 149.
- Adams, J. A., Stenson, H. H., & Humes, J. M. (1961). Monitoring of complex visual displays: II. Effects of visual load and response complexity on human vigilance. *Human Factors*, 3(4), 213.
- Bakan, P. (1955). Discrimination decrement as a function of time in a prolonged vigil. *Journal of Experimental Psychology*, 50(6), 387.
- Bevan, W., Avant, L. L., & Lankford, H. G. (1967). Influence of Interpolated Periods of Activity and Inactivity Upon the Vigilance Decrement. *Journal of Applied Psychology*, 51(4), 352-356.
- Dittmar, M. L., Warm, J. S., Dember, W. N., & Ricks, D. F. (1993). Sex differences in vigilance performance and perceived workload. *Journal of General Psychology*, 120(3), 309.
- Fisk, A. D., & Schneider, W. (1981). Control and automatic processing during tasks requiring sustained attention: A new approach to vigilance. *Human Factors*, 23(6), 737.
- Frankmann, J. P., & Adams, J. A. (1962). Theories of Vigilance. *Psychological Bulletin*, 59(4), 257-272.
- Jerison, H. J., & Pickett, R. M. (1963). Vigilance: A review and re-evaluation. *Human Factors*, 5(3), 211.
- Mackworth, N. H. (1948). The breakdown of vigilance during prolonged visual search. *The Quarterly Journal of Experimental Psychology*, 1, 6.
- McIntosh, E. G. (2006). Sex differences in boredom proneness. *Psychological Reports*, 98(3), 625-626.
- Pattyn, N., & Soetens, E. Endogenous and Exogenous Attention in Vigilance Tasks.
- Prinzel, L. J., & Freeman, F. G. (1997). Task-specific sex differences in vigilance performance: Subjective workload and boredom. *Perceptual and Motor Skills*, 85(3), 1195-1202.
- Scott, T. H. (1957). Literature review of the intellectual effects of perceptual isolation. *Dept. Nat. Def. Rep.*(66), iii.
- Smith, R. L., Lucaccini, L. F., Groth, H., & Lyman, J. (1966). Effects of anticipatory alerting signals and a compatible secondary task on vigilance performance. *Journal of Applied Psychology*, 50(3), 240.
- Tolin, P., & Fisher, P. G. (1974). Sex-Differences and Effects of Irrelevant Auditory-Stimulation on Performance of a Visual Vigilance Task. *Perceptual and Motor Skills*, 39(3), 1255-1262.

Waag, W. L., Halcomb, C. G., & Tyler, D. M. (1973). Sex differences in monitoring performance. *Journal of Applied Psychology*, 58(2), 272.

Williges, R. C. (1969). Within-session Criterion Changes Compared to an Ideal Observer Criterion in a Visual Monitoring Task. *Journal of Experimental Psychology*, 81(1), 61-66.