

# The Pale Shades of Emotion: A Signal Detection Theory Analysis of the Emotional Stroop Task

Boaz M. Ben-David<sup>1,2,3</sup>, Eran Chajut<sup>4</sup>, Daniel Algom<sup>5</sup>

<sup>1</sup>Toronto Rehabilitation Institute, Toronto, Canada

<sup>2</sup>Department of Speech-Language Pathology, University of Toronto, Toronto, Canada

<sup>3</sup>Department of Psychology, University of Toronto Mississauga, Mississauga, Canada

<sup>4</sup>Department of Education and Psychology, The Open University of Israel, Ra'anana, Israel

<sup>5</sup>Department of Psychology, Tel-Aviv University, Tel-Aviv, Israel

Email: boaz.ben.david@utoronto.ca

Received April 12<sup>th</sup>, 2012; revised May 1<sup>st</sup>, 2012; accepted May 29<sup>th</sup>, 2012

In the emotional Stroop effect (ESE), people are slower to name the ink color of negative, emotion-laden words than that of neutral words. Two accounts have been suggested for the ESE, relating it to either deficient attention to color or to temporary disruption of action in the face of threat. Is the ESE driven by a threat-produced change in perception, or is it a strategic bias in responding? In a pioneer import of Signal Detection Theory to this realm, threat was found to diminish the psychological distance ( $d'$ ) between the ink colors, but it did not impact response bias ( $\beta$ ). The results indicate that the ESE derives from a deep perceptual change engendered by the negative stimuli and not from changes in the criterion for responding. These results constrain future theorizing in the domain of emotion-produced changes in behavior, and provide some support for the threat account of attention under emotion.

**Keywords:** Emotional Stroop Effect; Attention; Threat; Signal Detection Theory; Emotion

## Introduction

Can people focus on an attribute of the stimulus when another attribute is laden with emotion or directly threatening? There is a voluminous literature concerned with this question employing mainly the paradigm known as the emotional Stroop task (e.g., Algom, Chajut, & Lev, 2004; Algom, Zakay, Monar, & Chajut, 2009; Chajut, Mama, Levi, & Algom, 2010; McKenna & Sharma, 1995, 2004; Watts, McKenna, Sharrock, & Trezise, 1986; Williams, Mathews, & MacLeod, 1996). Presented with words in color, it takes people longer to name the ink color of emotion or threat words than that of neutral words, the emotional Stroop effect (ESE). The ESE is a robust phenomenon, yet its source is debated in the literature.

According to the *attention account* of the ESE (e.g., Williams et al., 1996), the disproportionate amount of resources drawn by the emotion words takes a toll on naming their ink color. It is widely documented that threatening stimuli grab attention even when the negative information comprises a word presented in the shielded environment of the laboratory. The experimental task calls for naming ink colors, yet people cannot avoid reading the carrier words, emotional and neutral. It is the extra amount of attention captured by the former that generates the ESE. According to the *threat approach* to the ESE, the menacing content of the word engenders a temporary freeze on all ongoing activity. The interruption, if for a fraction of a second, sustains the prioritizing of resources for efficient action in the face of (potential) threat (Öhman, Flykt, & Esteves, 2001). The ESE thus reflects the activity of a general-purpose defense mechanism that momentarily freezes all activity that is not directly related to the threat (Algom et al., 2004; Fox, Russo, Bowles, & Dutton, 2001).

Why are people sluggish to respond to a feature (color) of emotion stimuli? Are they distracted by the emotion word or are they paralyzed by its threatening content? Indeed, it may be the case that both the attention and the threat theories describe the same phenomenon from different points of vantage (Frings, Englert, Wentura, & Bermeitinger, 2010). The goal of this study was to ponder further the nature of the change wrought about by responding to emotion stimuli. Does the processing of the threatening content of the word creates a sensory change, or does it bias the response? In the current study, we elucidate this issue by harnessing the potent tool of the theory of signal detectability (TSD; Macmillan & Creelman, 2005; cf. Marks & Algom, 1998).

## Applying Detection Theory to the Study of the ESE

Detection theory is a general approach to measuring decision under uncertainty. In a typical experiment, one of two possible stimuli (designated, respectively, as signal-and-noise and noise) is presented and the observer must report on each trial the identity of the stimulus (= whether or not the signal was presented). The decision is affected by the *sensitivity* to the physical difference separating the stimuli and by the criterion or *response bias* adopted, favoring one stimulus over the other in responding. Detection theory provides distinct procedures to calculate sensitivity and response bias separately (where each index can remain invariant in the face of changes in the other). The most widely used measure to tap sensitivity is  $d'$ , whereas the most popular measure to mark bias is  $\beta$ . We calculated  $d'$  and  $\beta$  for classification of the ink colors for each participant in each condition of the present study.

In the current adaptation, a word appeared in one of two col-

ors. The participants performed in the routine ESE task, namely, they identified the ink color of each word. The carrier words—irrelevant to the task at hand—were negative in one block, but neutral in another block. A voluminous literature suggests that the valence of the carrier words modifies performance—the ESE. What we do not know is the source of this change in performance under emotion or threat. Is it rooted in sensory modification or in cognitive adjustment? The former would be expressed in a difference in the values of  $d'$  across the emotion and neutral blocks, the latter in a difference in the pertinent values of  $\beta$  across the two blocks.

### Does Emotion Affect Stimulus Discriminability or Response Bias?

TSD permits to disentangle effects of perception from those of response bias—a critical step for understanding behavior under emotion. Why should threat have an impact on the perception of colors or on the strategy for responding? First, consider the sensory index,  $d'$ . If resources are prioritized to deal with threat, as discussed in the threat account, it comes at the expense of poorer perception of other attributes, notably that of the pertinent colors. As a result, perceptual sensitivity for color can be affected under emotion or threat, so that the ink colors become psychologically closer to one another. Longer color naming under emotion (=ESE) ensues.

Consider next the  $\beta$  index, the criterion espoused for reporting. Fast and automatic processing of threat can modify the disposition to respond affirmatively. For example, Windmann and Krüger (1988) found a bias-induced change in recognition of emotion words. The authors conclude that “aversive stimulus valence had affected subjects’ willingness to risk false-positive responses via automatic and unconscious influences” (p. 625). More recently, Windmann and Chmielewski (2008) found that changing the response from affirmative to negative (respond “yes” or “no” for an “old” item) sufficed to reverse the bias for detection of threat words (in a memory task). Similar results were obtained in detection of negative (and neutral) pictures (e.g., Wiens, Piera, Golkar, & Öhman, 2008) and faces (Westermann & Lincoln, 2010).

Clearly, importing the TSD into the realm of emotion opens up new avenues for research (in this case, for studying the ESE). Can it also provide a means for theoretical resolution? It is not fully clear how to derive the respective predictions by the attention and the threat accounts with respect to the expected results of a TSD analysis. Yet, it appears that a change in the sensory index ( $d'$ ) of TSD across emotion and neutral stimuli is more closely linked with the threat theory, as this theory entails a deep perceptual modification under threat. A change in the  $\beta$  index of TSD, the criterion espoused for reporting, may be more consistent with the attention theory that does not entail a genuine perceptual modification in emotion. We admit that these predictions are rather preliminary and are best considered as working hypotheses at this point.

### Accuracy-Based ESE

Finally, a notable feature of TSD analyses is that they are based on accuracy or error, not on RT. As a consequence, we gauged the ESE for accuracy for the first time in the emotion literature (but see Zeelenberg, Wagenmakers, & Rotteveel, 2006, for a non-ESE study with accuracy). The problem, of

course, is the minuscule rates of errors observed in virtually all studies of the ESE. How does one generate sufficiently high rates for error to permit the TSD analysis? Our tactic was two-fold. First, we imposed a severe time window on responding (cf. Lindsay & Jacoby, 1994). This encouraged very speedy responding at the expense of accuracy. Second, we also used imperfectly discriminable print colors (red and orange). The two manipulations conspired to produce sufficiently high rates of error, allowing an accuracy-sustained look at the ESE. In addition, our task was fashioned so that one font color called for an affirmation response and the other for a rejection.

## Method

### Participants

Twenty-eight young adults (between 18 and 26 years of age), Tel-Aviv University undergraduate students, participated in the experiment against course credit. All were native speakers of Hebrew and all had normal or corrected to normal vision (and color-vision) assessed by self-report

### Stimuli and Apparatus

There were 16 words of various clothing for neutral items and 16 words associated with terror for negative items (see Appendix). Note that both types of words were drawn from single well defined categories, a constraint adopted in order to rule out any effect of category. The words were equally familiar based on the average rating of an independent group of 40 Tel-Aviv University students. The students used a scale between 0 (unfamiliar) and 5 (very familiar) to assess 64 words. Included in the study were emotion and neutral words with an average rating between 3 and 4 (inclusive). The two sets of words were also matched in word length (for a review on the impact of these lexical characteristics, see Ben-David, Van Lieshout, & Leszcz, 2011).

The words appeared on the white background of a 17" color monitor (set to a resolution of 1024 × 768 pixels). They were presented at the center of the screen. However, in order to avoid adaptation and/or strategic responding (e.g., fixating on a small portion of the print to avoid reading the words when responding to the color), we introduced a trial-to-trial spatial uncertainty of 15 pixels around the target location. The words were printed in Arial (Hebrew font, size 48) placed within the invisible frame of a 118 × 40 pixels rectangle. Viewed from a distance of 60 cm, the word subtended 4.48° of visual angle in width and 1.52° in height. The words appeared in the prototypical colors of red or orange (a difference in hue of 15%, based on the software’s standards) rendering the classification difficult. Each word was preceded by a row of 7 Xs (black, Arial font size 48) serving as a mask. We used a loud tone (2000 Hz, 60 dB, 200 ms) to signal the end of the time window for responding (an error from the participant’s point of vantage). The tones were played over a pair of (Peerless) loudspeakers.

### Design

The words were presented singly for view. There were two separate blocks of trials, one entailing the emotion words, and the other entailing the neutral words (with order counterbalanced across participants). Within the block, each word appeared in each of the two ink colors six times, making for 192

experimental trials per block. The order of these stimuli was random and different for each participant. On a trial, the mask was presented at the center for 50 ms, followed by the word in color. The time window for responding was set at 500 ms. A response produced within this limit terminated the presentation and the mask starting the next trial appeared after 50 ms. However, if the participant failed to respond during the 500 ms time window, the presentation terminated with the “error” tone played aloud. Again, the next trial started after a 50 ms interval.

## Procedure

The participant decided whether the ink color was red (“yes”) or not red (“no”, i.e., orange) by pressing one of a pair of lateralized keys (A and L; key assignment was counterbalanced between participants). The participant was asked to do so within a 500 ms time frame. The participant was further informed that failure to respond within the time limit will be penalized by an unpleasant tone, and that the trial will be discarded as erroneous. Each block was preceded by 8 training trials. Unbeknownst to the participant, the first 10 trials in each experimental block were also considered as practice and discarded from the analysis.

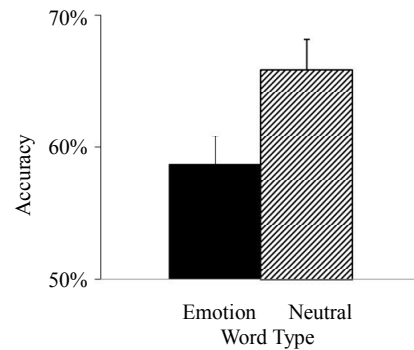
## Data Analysis

For purposes of the TSD analysis we considered only the responses made during the 500 ms interval (88% and 86% in the emotion and neutral blocks, respectively,  $t(27) = 1.8$ ,  $p = .08$ ). We defined the red color as signal-and-noise (hence, orange as noise). Consequently, the percentage of correct identification of red provided the rate of hits (H). In a complementary fashion, the percentage of incorrect identifications of orange (as red) provided the rate of false alarms (FA). Based on these rates, we calculated the sensory index of discriminability ( $d' = Z(H) - Z(FA)$ ) and the decisional index of criterion ( $\beta = \exp[0.5 * ((Z(FA))^2 - (Z(H))^2)]$ ) separately in each block.

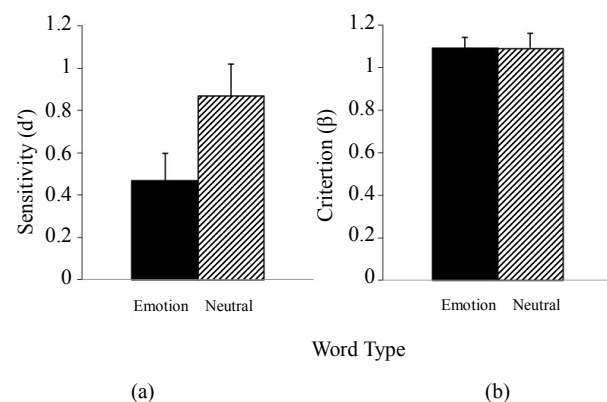
## Results

The accuracy of identifying color in the blocks with emotion and neutral words is presented in **Figure 1**. The first feature to notice is the unusually high rates of error committed (hovering at around 37% overall). We conclude that our manipulations were successful in generating sufficiently high percentage of errors to allow accuracy based analyses. Note, too, that **Figure 1** is the standard ESE plot with the notable exception that accuracy replaces RT on the ordinate. Under the inauspicious conditions of the experiment, our participants were more accurate to detect the ink color of a neutral word (65.9%) than that of an emotion word (58.6%). This difference favoring neutral words as carriers of color amounted to an ESE of 7.3% ( $t(27) = 6.7$ ,  $p < .001$ ).

Whence the disruption of performance with negative stimuli (= ESE)? First consider the data with respect to the sensitivity of color discrimination presented in the left-hand panel of **Figure 2**. The average  $d'$  in the block with neutral items (0.869) was almost twice that in the block with the emotion items (0.468;  $t(27) = 6.48$   $p < .001$ ). Clearly, our participants were more sensitive to the difference between the same ink colors when the carrier words were neutral than when the carrier words were emotional.



**Figure 1.** Percent correct for naming the color of emotion and neutral words under time pressure.



**Figure 2.** Results of the TSD analysis. Panel A: Average values of perceptual discrimination,  $d'$ , for ink color of emotion and neutral words. Panel B: Average values of response bias,  $\beta$ , for ink color of emotion and neutral words. Error bars represent one standard error around the mean.

Finally, consider the data with respect to response bias in the right-hand panel of **Figure 2**. We did not discern any consistent pattern of the values of  $\beta$  across the two blocks. The net result of this unsystematic variability was that the mean values in the two blocks were almost fully comparable (at around 1.09). We conclude that the response criterion adopted by the participants was not modified by stimulus valence. The difference in outcome between the  $d'$  and the  $\beta$  was further supported by the interaction of word valence and TSD measure ( $F(1,27) = 32.2$ ,  $p < .001$ ,  $MSE = 0.035$ ,  $\eta_p^2 = .55$ ).

## Discussion

This early foray into TSD based emotion research already yielded important results. They show that the content of the word exerts a deep sensory effect on the processing of its ink color. The presence of threat in the form of a negative or emotional word diminishes the psychological distance separating the ink colors of the carrier words. In general, people are desensitized to the differences between stimulus attributes that are unrelated to the threat attribute. In contrast, response bias is not altered by threat. The upshot is that the ESE derives from perceptual changes in the environment engendered by the presence of threat. The slowdown with emotion items is the toll exacted on performance by the (temporarily) dilapidated perception. In

other words, threat-related cost also entails poorer discrimination of stimulus attributes.

The results mandate the conclusion that the ESE is the outcome of an instinctive perceptual-motor reaction to threat. The presence of threat engenders a wholesale regrouping of the organism's resources in order to deal effectively with the situation. Resources are allocated to the threat attribute at the expense of other attributes. Perception becomes thus blunt for attributes other than the threat one. This finding is consistent with recent results from our laboratory showing that non-target attributes often do not undergo semantic processing in the presence of emotion or threat (Chajut, Schupak, & Algom, 2010). The findings are also consistent with those by Zeelenberg et al. (2006) and by Schupp, Junghofer, Weike, and Hamm (2003), which implicate a perceptual source for changes in performance under emotion. Taken together, these results provide further support for the threat account of the ESE, where the sluggish performance (otherwise termed "temporary freezing") in the face of threat is the cost of reshuffling of priorities.

Incidental support for this conclusion comes from a pair of recent studies on the classic Stroop effect that manipulated color (actually, the discriminability of the ink colors) in a direct fashion. Ben-David and Schneider (2009, 2010; see also Ben-David, Nguyen, & Van Lieshout, 2011) found that the content of the word exerted a greater influence on color naming (thereby generating a larger Stroop effect) when the colors were less salient. Within the framework of the ESE, the natural salience of a threat stimulus combines with poorer color perception (itself produced by the threat) to generate slower color naming.

In conclusion, let us issue a caveat with respect to the general impact of the current results. First, they must be replicated and extended to further stimuli, populations, and areas of stress and anxiety. Second, the null-effect with respect to response bias derives from large unexplained variability. It would be salutary to replicate it in the face of reduced variance. Nevertheless, our data constrain future theories of the ESE, challenging any a strategy-driven source for the effect.

### Acknowledgements

The first author was partially supported by a grant from the Ontario Neurotrauma Foundation (2008-ABI-PDF-659). We wish to thank Noa Calderon for her assistance in collecting the data.

### REFERENCES

- Algom, D., Chajut, E., & Lev, S. (2004). A rational look at the emotional Stroop phenomenon: a generic slowdown, not a Stroop effect. *Journal of Experimental Psychology: General*, *133*, 323-338. doi:10.1037/0096-3445.133.3.323
- Algom, D., Zakay, D., Monar, O., & Chajut, E. (2009). Wheel chairs and arm chairs: A novel experimental design for the emotional Stroop effect. *Cognition and Emotion*, *23*, 1552-1564. doi:10.1080/02699930802490243
- Ben-David, B. M., Nguyen, L. L. T., & Van Lieshout, P. H. H. M. (2011). Stroop effects in persons with traumatic brain injury: Selective attention, speed of processing or color-vision? *Journal of the International Neuropsychological Society*, *17*, 354-363. doi:10.1017/S135561771000175X
- Ben-David, B. M., & Schneider, B. A. (2009). A sensory origin for color-word Stroop effects in aging: A Meta-Analysis. *Aging Neuropsychology and Cognition*, *16*, 505-534. doi:10.1080/13825580902855862
- Ben-David, B. M., & Schneider, B. A. (2010). A sensory origin for color-word Stroop effects in aging: Simulating age-related changes in color-vision mimics age-related changes in Stroop. *Aging Neuropsychology and Cognition*, *17*, 730-746. doi:10.1080/13825585.2010.510553
- Ben-David, B. M., Van Lieshout, P. H. H. M., & Leszcz, T. (2011). A resource of validated affective and neutral sentences to assess identification of emotion in spoken language after a brain injury. *Brain Injury*, *25*, 206-220. doi:10.3109/02699052.2010.536197
- Chajut, E., Mama, Y., Levi, L., & Algom, D. (2010). Avoiding the approach trap: A response bias theory of the emotional Stroop effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1567-1572. doi:10.1037/a0020710
- Chajut, E., Schupak, A., & Algom, D. (2010). Emotional dilution of the Stroop effect: A new tool for assessing attention under emotion. *Emotion*, *10*, 944-948. doi:10.1037/a0020172
- Fox, E., Russo, R., Bowles, R., & Dutton, K. (2001). Do threatening stimuli draw or hold visual attention in subclinical anxiety? *Journal of Experimental Psychology: General*, *130*, 681-700. doi:10.1037//0096-3445.130.4.681
- Frings, C., Englert, J., Wentura, D., & Bermeitinger, C. (2010). Decomposing the emotional Stroop effect. *Quarterly Journal of Experimental Psychology*, *63*, 42-49. doi:10.1080/17470210903156594
- Lindsay, D. S., & Jacoby, L. L. (1994). Stroop process dissociations: the relationship between facilitation and interference. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 219-234. doi:10.1037/0096-1523.20.2.219
- Macmillan, N. A., & Creelman, C. (2005). *Detection theory*. Mahwah, NJ: Lawrence Erlbaum.
- Marks, L. E., & Algom, D. (1998). Psychophysical scaling. In M. H. Birnbaum (Eds.), *Measurement, judgment, and decision making. Handbook of perception and cognition* (pp. 81-178). San Diego, CA: Academic Press.
- McKenna, F. P., & Sharma, D. (1995). Intrusive cognitions: An investigation of the emotional Stroop task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1595-1607. doi:10.1037/0278-7393.21.6.1595
- McKenna, F. P., & Sharma, D. (2004). Reversing the emotional Stroop effect reveals that it is not what it seems: The role of fast and slow components. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 382-392. doi:10.1037/0278-7393.30.2.382
- Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *Journal of Experimental Psychology: General*, *130*, 466-478. doi:10.1037//0096-3445.130.3.466
- Schupp, H. T., Junghofer, M., Weike, A. I., & Hamm, A. O. (2003). Emotional facilitation of sensory processing in the visual cortex. *Psychological Science*, *14*, 7-13. doi:10.1111/1467-9280.01411
- Watts, F. N., McKenna, F. P., Sharrock, R., & Trezise, L. (1986). Colour naming of phobia-related words. *British Journal of Psychology*, *77*, 97-108. doi:10.1111/j.2044-8295.1986.tb01985.x
- Westermann, S., & Lincoln, T. M. (2010). Using signal detection theory to test the impact of negative emotion on sub-clinical paranoia. *Journal of Behavior Therapy and Experimental Psychiatry*, *41*, 96-101. doi:10.1016/j.jbtep.2009.10.007
- Wiens, S., Peira, N., Golkar, A., & Öhman, A. (2008). Recognizing masked threat: Fear betrays, but disgust you can trust. *Emotion*, *8*, 810-819. doi:10.1037/a0013731
- Williams, J. M. G., Mathews, A., & MacLeod, C. (1996). The emotional Stroop task and psychopathology. *Psychological Bulletin*, *120*, 3-24. doi:10.1037/0033-2909.120.1.3
- Windmann, S., & Chmielewski, A. (2008). Emotion-induced modulation of recognition memory decisions in a Go/NoGo task: Response bias or memory bias? *Cognition and Emotion*, *22*, 761-776. doi:10.1080/02699930701507899
- Windmann, S., & Krüger, T. (1998). Subconscious detection of threat as reflected by an enhanced response bias. *Consciousness and Cognition*, *7*, 603-633. doi:10.1006/ccog.1998.0337
- Zeelenberg, R., Wagenmakers, E. J., & Rotteveel, M. (2006). The impact of emotion on perception: Bias or enhanced processing? *Psychological Science*, *17*, 287-291. doi:10.1111/j.1467-9280.2006.01700.x

## Appendix

### List of the Words Used in the Study (Translated from Hebrew)

*Neutral words:* Scarf, Hat, Glove, Shirt, Pants, Coat, Sweater, Umbrella, Vest, Boot, Dress, Skirt, Tie, Sandal, Sock and Shoe.

*Emotion words:* Murder, Suicide, Terrorist, Danger, Atrocious, Extermination, Death, War, Suffocation, Injured Person (a single word in Hebrew), Horror, Poison, Terrorist Attack (a

single word in Hebrew, lexically unrelated to Terrorist), Burn, Mucus, and Scare.

### The Original Hebrew Words

*Neutral words:* סוודר, כפפה, חולצה, מכנסים, מעיל, טוודר, מטרייה, צעיף, כובע, גרבי, סנדל, מגף, שמלה, הצאית, עניבה, סנדל, נעל, גרב, סרב, מגף, שמלה, הצאית, עניבה, סנדל,

*Emotion words:* רצח, מתאבד, מחבל, סכנה, זוועה, השמדה, מוות, מלחמה, חנק, פצוע, אימה, רעל, פיגוע, כווייה, מוגלה, בהלה