The Effect of Exposure to Violence on Trustworthiness Perception

Katherine Hong

Yale University
Senior Thesis, Cognitive Science
Advised by Arielle Baskin-Sommers
Spring 2019

Abstract

Individuals exposed to violence, especially during childhood, show negative mental and physical health effects, impaired empathetic responses and impulse control, and generally more aggressive, anti-social behaviors. However, there is limited research on how exposure to violence interferes in specific pro-social perception mechanisms to result in aggressive behavior. Trustworthiness perception is an important mechanism that affects social decision making, but has not been directly studied with an eye towards exposure to violence. Participants in this study completed both explicit and implicit trustworthiness rating tasks, while electroencephalography was recorded. Results show that individuals exposed to violence demonstrate increased late positive potential (LPP; a measure of motivated affective elaboration) when making explicit trustworthiness ratings about untrustworthy and neutral faces. When making implicit trustworthiness judgments, individuals exposed to violence showed increased P300 amplitude (a measure of event-related potential) across all face types. There were no significant differences in behavior within tasks for individuals exposed to violence, indicating that exposure to violence does not impair trustworthiness perception, but changes the amount of neural resources necessary to complete the task. This gives insight into how exposure to violence affects social decision making.

Introduction

Imagine a neighborhood without trust. Perhaps there is rampant crime, visible disorder, and limited resources. A large feature of these disadvantaged neighborhoods is the prevalence of violence. American history of racialized and urbanized poverty has created regions of concentrated disadvantage, characterized by high neighborhood levels of violence (Anderson, 1994). However, how are these disadvantages explicitly measured?

Exposure to violence (ETV) is an important community factor that demonstrates the experience of individuals within broader social trends. Children who grow up in disordered neighborhoods, both suburban and urban, are likely to report high levels of both witnessing and directly experiencing violence (Campbell & Schwarz, 1997; Schubiner, Scott, & Tzelepis, 1993). Exposure to violence in childhood is linked to behavioral issues in adolescence (Guerra, Rowell Huesmann, & Spindler, 2003; Mrug & Windle, 2008) as well as antisocial behaviors that persist into and throughout adulthood (Stouthamer-Loeber, Loeber, Wei, Farrington & Wikström, 2002). Beyond behavior, these adverse experiences have negative health effects, both mental and physical, upon residents (Bailey et al., 2005; Baskin & Sommers, 2014; Fowler et al., 2009). Deficits in neural features such as IQ, reading, empathetic responses, and impulse control point to the effect that ETV has upon the brain itself (Delaney-Black et al., 2002; Guo et al., 2013; Monahan, King, Shulman, Cauffman, & Chassin, 2015). Although there is evidence that ETV is correlated with deficits in social processes, there is a dearth of research surrounding ETV's effect on certain pro-social mechanisms, specifically trustworthiness perception.

Trustworthiness is an important social signal that is the foundation of meaningful relationships. Research shows that, even in the absence of information about the other person's behavior or emotional state, individuals form stable trustworthiness judgments about another person based solely on their facial features. Judgments of trustworthiness mainly rely on four features: the inner eyebrows, cheekbones, chins, and nose sellion (Oosterhof & Todorov, 2008; Todorov, Baron, & Oosterhof, 2008). A trustworthy face is characterized by higher inner eyebrows, pronounced cheekbones, wide chins, and shallow nose sellion. Untrustworthy faces are characterized by the opposite direction of these features. Judgments about trustworthiness are formed within just 100 ms of exposure to the face (Willis & Todorov, 2006). Untrustworthy faces evoke larger brain responses by 300 ms (i.e., late positive potential (LPP), a measure of motivated elaborative processing; Foti, Hajcak, & Dien, 2009; Hajcak, Dunning, & Foti, 2009). LPP as a measure of motivated elaborative processing has been studied as an important step in visual processing, especially along aspects of emotional valence and arousal (Lang, Bradley, & Cuthbert, 1997). LPP and other kinds of event related potentials (ERP) indicate how many neural resources are allotted to completing a perceptual task.

Research demonstrates larger LPP to subjectively (Lischke, Junge, Hamm, & Weymar, 2017; Marzi, Righi, Ottonello, Cincotta, & Viggiano, 2014) and pre-classified (Yang, Qi, Ding, & Song, 2011) untrustworthy, compared with neutral and trustworthy, faces. People use information such as facial appearance to predict the trustworthiness of an unknown person (Bonnefon, Hopensitz, & De Nys, 2013; Willis & Todorov, 2006) and decide how to engage with that person.

To examine the relationship among exposure to violence, facial trustworthiness, and behavior, we used a series of tasks paired with electroencephalography (EEG) in a community sample enriched for aggressive behavior. The study was meant to examine both the neural mechanism and then make a connection to real world behavior. This research is building off the aforementioned work studying trustworthiness and the long-term effects of ETV, but as of yet there is a gap in research at the intersection of the two. These analyses were meant to be exploratory, and to help determine where further research should be directed. First, we were interested in the relationship between exposure to violence and both behavioral and neural responses to facial trustworthiness. Participants rated faces manipulated on trustworthiness as "untrustworthy," "trustworthy," or "neutral." Second, we were interested in how exposure to violence and subjective perceptions of trustworthiness were related to self-reported aggressive behavior. Examining how exposure to violence affects trustworthiness perception from a neural perspective can give insight into how these adverse experiences directly lead to antisocial behavior.

Two questionnaires were administered to participants to give measures of ETV and aggression, based in real world experiences and behavior. Participants completed two tasks, one where trustworthiness of novel faces was explicitly rated, and one where trustworthiness perception was implicit. The first task, the "Trustworthiness Rating Task," gives information about ETV's effect on explicit trustworthiness judgments, both in behavioral outcome (rating) and neural activity (LPP). This task should give an answer about whether ETV changes the actual trustworthiness judgments given to faces, as well as any neural difference during judgements. The second, the "Implicit Approach Avoidance Task," measured implicit trustworthiness perception through task performance, as well as the neural activity involved in implicit judgment. This data is important to determining whether ETV causes different performance when implicitly judging trustworthiness while focusing on a different goal, as well as what neural effects ETV may have on this process. Finally, the aggression and ETV measure were collected to examine how ETV and trustworthiness perceptions interact to produce aggressive behavior, the second research question. These measures and two tasks build upon previous research in ETV and trustworthiness perception, and seek to connect the neural mechanisms to real world behavior.

Methods

Participants

We used a targeted recruitment approach in a high-crime community to enrich our sample for aggressive behavior. Sixty-six aged 18 to 75 (*M*=38, *SD*=12) were recruited from the community through flyers calling for individuals who engage in risk-taking behavior (e.g., crime, bullying, substance use, gambling, impulsive behavior) in New Haven County, Connecticut.

A prescreen phone interview and in person clinical assessment were used to exclude individuals who were younger than 18 or over 75, had performed below a fourth-grade level of a standardized measure of reading (WRAT-III; Wilkinson, 1993), who scored below 70 on a brief measure of IQ (Shipley; Zachary, 1986), who had diagnoses of schizophrenia, bipolar disorder, or psychosis not otherwise specified, or who had a history of medical problems (e.g., uncorrectable auditory or visual deficits, seizures, head injury with loss of consciousness greater than 30 minutes, color blindness) that may impact their comprehension of materials and performance on the task. All participants provided written informed consent and experimental procedures were approved by the Yale University Human Investigation Committee. Participants were paid $10/hour for completion of the self-report measure and experimental tasks.

Measures

The Risky Impulsive Self-Destructive Questionnaire (RISQ) is a 38-item self-report questionnaire that measures risky, impulsive, and self-destructive behaviors in 8 domains (aggression, self-harm, gambling, impulsive spending/driving, impulsive eating, risky sex, illegal behavior, and alcohol use). For each behavior, respondents note the number of times they have engaged in the behavior in their lifetime, how many times in the past month, and how old they were when they first started engaging in the behavior. Additionally, the respondents indicate if there were any consequences (e.g., legal, relationship, financial) as a result of their behavior (dichotomous: yes/no). Finally, for each behavior respondents indicate how strongly they agree with statements that assess their motivation (distress or pleasure) for engaging in the behavior (0=Strongly disagree; 1=Somewhat disagree; 2=Equally disagree/agree; 3=Somewhat agree; 4=Strongly Agree). The RISQ has been found to be a valid measure of antisocial and externalizing behaviors (Sadeh & Baskin-Sommers, 2016).

The Exposure to Violence scale (ETV) is a 13-item self-report measure of experienced lifetime violence, either through personal victimization or witnessing of an event. Participants are asked to circle *Yes* or *No* for each item. Examples of items include “Have you been hit, slapped, punched, or beaten up?” and “Have you seen someone else get attacked with a weapon, like a knife or bat?” If *Yes* is endorsed, participants must indicate the number of times they have experienced this situation in their lifetime. Total Scores range from 0-13, with higher scores indicating a greater lifetime exposure to violent situations. This has been found to be a reliable measure of exposure to personal and witnessed violence (Selner-O’Hagan, Kindlon, Buka, Raudenbush, & Earls, 1998).

Experimental Tasks

Apparatus. Presentation of all stimuli and measurement of behavioral responses was controlled using the Pyschtoolbox extension in MATLAB. EEG recording was controlled by a MATLAB script and Neuroscan Synamps amplifiers and acquisition software (Compumedics, North Carolina). All tasks were presented on a Ben-Q 27-inch high performance LED gaming monitor. Participants' eyes were about at a distance of 75 cm from the screen. Participants registered their responses using a button box.

Stimuli. Stimuli were obtained from a set of computer-generated faces manipulated on a dimension of trustworthiness based on methods described by Oosterhof and Todorov (2008). The faces were originally generated using FaceGen Modeller 3.2 (Singular Inversions, 2007), and are all male, bald, Caucasian, and front facing with direct gaze. The set contains a total of 100 unique identities varied on three levels of trustworthiness, for a total of 300 faces. The three levels of trustworthiness represented less trustworthy (-3 SD), neutral (0 SD), and more trustworthy (+3 SD) versions of each facial identity (see Figure 1). This dataset has been used successfully in previous research evaluating the trustworthiness of faces and prior EEG studies investigating evaluation of facial trustworthiness (Marzi et al., 2014; Oosterhof & Todorov, 2008; Yang et al., 2011). Furthermore, use of this dataset allowed the ratings of trustworthiness made by aggressive individuals to be compared against broader consensus ratings.

Figure 1. Example facial identity varied on three levels of trustworthiness from left to right: untrustworthy (-3 SD), neutral (0), and trustworthy (+3 SD).

Trustworthiness Rating Task

Participants were instructed to rate the trustworthiness of the face that appeared on the screen. It was emphasized that the face would be presented briefly and that participants should go with their "gut reaction" when rating the face. When prompted by a ratings screen, participants rated the presented face by pressing either "trustworthy", "neutral", or "untrustworthy" labeled buttons on a button box with the index finger of the dominant hand.

In total, participants completed 180 trials comprised of 60 randomly selected identities that varied on 3 levels of trustworthiness. The trustworthiness ratings task consisted of 6 blocks of 30 trials, with the order of the face stimuli randomized such that the same facial identity was not presented more than once in a block in order to control for repetition effects. Each trial consisted of a white fixation cross presented on a black background for 500 ms, followed by presentation of a target face for 750 ms, and finally an instruction screen: "Rate the face on trustworthiness." (See Figure 2). Participants had up to 2250 ms to respond with a button press.

If the participant did not respond within that time, the task automatically advanced to the next trial. The intertrial interval was variable between 1000 to 2000 ms.

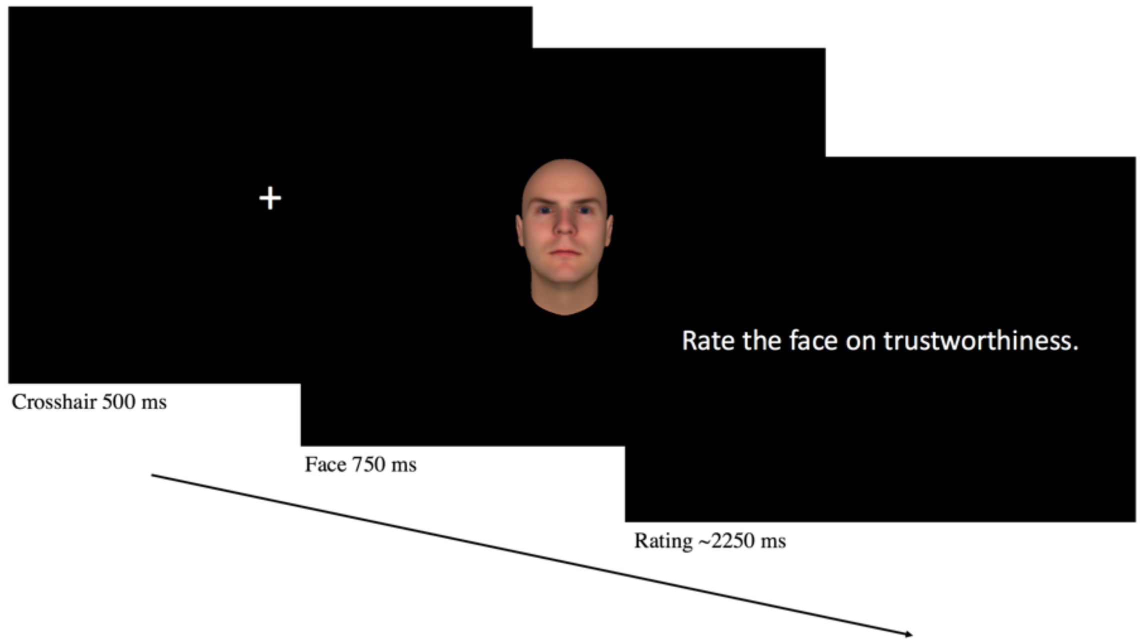

Figure 2. Sample trial from ratings task. Trials consisted of a 500 ms crosshair, a 750 ms presentation of a target face, and a screen prompting: rate this face on trustworthiness. Participants had up to 2250 ms to make a rating, otherwise the trial automatically advanced.

Implicit Approach-Avoidance Task

In the *Implicit Approach-Avoidance Task*, participants were instructed to imagine being a manikin that appeared on the screen. Participants were told that the goal was to move the manikin, or “yourself,” as quickly as possible up or down on the screen depending on the color of a box surrounding a face. Participants were to move the manikin by pressing one of two horizontal buttons with the index finger of their dominant hand.

The faces in this task were randomly selected from the participant’s subjective ratings in the *Trustworthiness Rating Task*; 30 faces rated as trustworthy, 30 rated as neutral, and 30 rated as untrustworthy were selected. If the participant rated fewer than thirty faces for any of the three categories, stimuli categorized based on Oosterhof and Todorov's original model (2008) were supplemented from identities that had not been utilized in the *Trustworthiness Ratings Task*.

The *Implicit Approach-Avoidance Task* consisted of 3 blocks of 30 trials. The combination of initial position of the manikin, the color box, and the level of trustworthiness of the face yielded six types of trials: approach trustworthy, avoid trustworthy, approach untrustworthy, avoid untrustworthy, approach neutral, and avoid neutral faces. Each participant viewed 15 trials of each trial type. Each trial began with a 500 ms presentation of a white fixation cross on a black background, followed by the manikin appearing either in the top or bottom half of the screen for 750 ms. A face surrounded by either a blue or a purple box then appeared in the middle of screen, and participants were told to move the manikin downward by pressing the down button if the color box was blue and upward by pressing the up button if the color box was purple. Depending on the initial position of the manikin and the color of the box, the manikin would then move to the top or bottom edge of the screen or next to the face for 500 ms. If the participant responded incorrectly, (i.e., pressed the down key for a purple color box), an error message would appear for 500 ms. Participants had up to 2250 ms to respond with a button press. If the participant did not respond within that time, the experiment automatically advanced to the next trial. The intertrial interval was variable between 1000-2000 ms (see Figure 3).

The combination of initial presentation of the manikin in the top or bottom half of the screen, color of the box surrounding the face, and category of trustworthiness of face was counterbalanced across trials. The mapping between color and upward versus downward response was kept consistent within participant, but between participants, this mapping was counterbalanced.

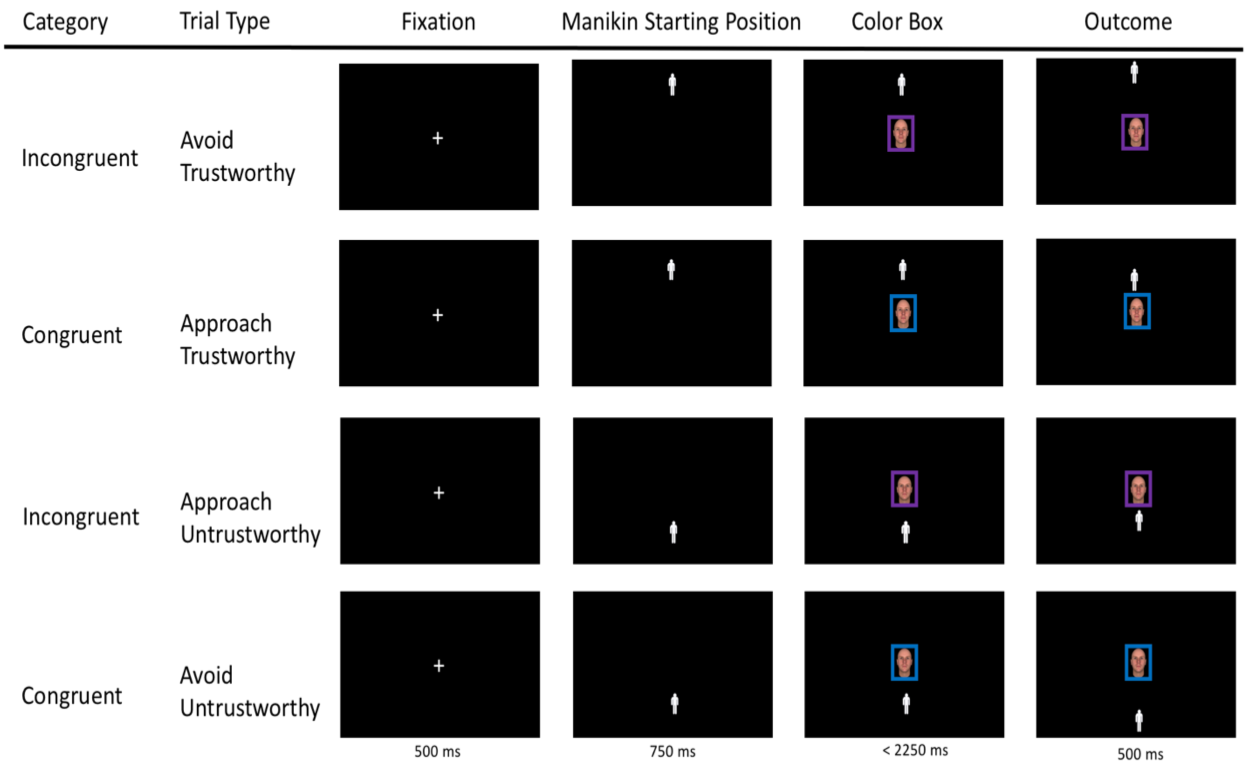

Figure 3. Illustration of the possible outcomes in one counterbalance for an untrustworthy face. Participants were instructed to move the manikin downward when a blue box surrounding a face appeared and to move the manikin upward when a purple box surrounding a face appeared. Initial position of the manikin in the upper or lower half of the screen caused upward or downward movement to imply approaching or avoiding the face. The mapping between color and upward versus downward response was counterbalanced across participants.

Procedure

For all tasks, participants sat in a comfortable chair in a dimly lit room facing the computer monitor. Before beginning each task, the experimenter read instructions aloud to the participants, and participants were given the opportunity to ask the experimenter for clarification before they initiated the tasks. Prior to beginning the *Implicit Approach-Avoidance Tasks*, participants were requested to restate the mappings of color versus upward and downward response or category of trustworthiness of face and upward versus downward response. If participants restated the incorrect mapping, the experimenter reminded the participant of the correct mapping, and the participant was asked to restate the correct mapping again. All participants completed tasks in the following order: *Trustworthiness Ratings Task*, *Implicit Approach-Avoidance Task*, *Explicit Approach-Avoidance Task* (not used in this paper), and *Control Task*.

Results

Explicit Trustworthiness Rating Task

Ratings. On average, participants rated 61 (*SD*= 26) faces as neutral, 47 (*SD*=26) faces as trustworthy, and 62 (*SD* = 24) faces as untrustworthy. To examine whether exposure to violence had any effect on trustworthiness rating, a linear regression was run. Results showed no significant effect of total ETV score on the faces that participants rated as trustworthy, $F(1, 64) = .00$, $p = .975$. To examine if ETV had an effect on faces that participants rated as untrustworthy, another linear regression was run, showing no significant effect, $F(1, 62) = .91$, $p = .343$.

In case ETV would affect congruency with Todorov's original model, a linear regression was conducted to examine the effect of ETV on participants' trustworthiness ratings congruence. The results showed no significant effect of ETV on how congruently participants rated faces as trustworthy, $F(1, 64) = .08$, $p = .738$. As not every participant saw every face, incongruency of trustworthiness ratings was not the exact inverse of the trustworthiness ratings congruency, so another ANOVA was conducted to examine the effect of ETV on participants' incongruence with the original model. This showed no significant results, $F(1, 64) = .12$, $p = .735$.

Another regression looked at the effect of ETV on the congruence of faces that participants rated as untrustworthy, demonstrating no significant effect, $F(1, 64) = .07$, $p = .789$. Again, incongruency in untrustworthiness ratings was not an exact inverse of congruency because of the variation in faces that participants saw. Results from the regression examining ETV and untrustworthiness incongruency showed no significant effect, $F(1, 64) = 1.01$, $p = .320$.

Neural Reactivity to Faces. To examine basic task effects, we entered mean LPP amplitude in a general linear model (GLM) with face type (neutral, trustworthy, or untrustworthy) as a within-subjects factor. Given that LPP amplitude is greater to affective stimuli, planned simple interaction contrasts of trustworthy vs. neutral and untrustworthy vs. neutral faces were conducted. Results showed no significant within-subjects effect of face type and ETV on LPP mean amplitude, $F(2, 126) = 1.46$, $p = .237$.

However, between subjects results demonstrated a significant main interaction between total ETV score with LPP mean amplitude, $F(1, 63) = 5.38$, $p = .024$. Follow-up simple effects tests showed that exposure to violence predicted greater LPP amplitude to neutral faces ($B = .237$ μV, $SE = .12$, $p = .043$, 95% CI = .01 to .47) and untrustworthy faces ($B = .32$, $SE = .12$, $p = .008$, 95% CI = .09 to .56), but was unrelated to LPP amplitude to trustworthy faces ($B = -.18$, $p = .126$).

Implicit Approach Avoidance Task

Task Effects. The General Linear Model showed no significant interaction between level of trustworthiness and face type $F = .10$, $p = .908$.

ETV. Results showed no significant interaction between exposure to violence, face type, and trial type, $F = .10$, $p = .905$.

To examine the neural mechanism, another GLM was run looking at P300 implicit stimulus evaluation. Results showed no significant within-subjects effect of face type and ETV on P300 mean amplitude, $F(1, 64) = .221$, $p = .640$.

Between subjects results demonstrated a significant interaction between total ETV score with P300 mean amplitude, $F(1, 64) = 5.47$, $p = .023$. Follow-up simple effects tests showed that exposure to violence predicted greater P300 amplitude to both trustworthy ($B = .28$ μV, $SE = .13$, $p = .028$, 95% CI = .03 to .54) and untrustworthy faces ($B = .33$ μV, $SE = .15$, $p = .033$, 95% CI = .03 to .63). Across all face types, ETV had a significant interaction with ERP.

Real World Behavior

Next, we examined P300 as mediator between ETV and self-reported aggression, measured from the RISQ. The moderation (interaction) model showed a significant main effect of ETV alone on aggressive behavior ($t = 6.10$, $p < .001$) and a non-significant main effect of P300 alone on RISQ aggression score ($t = 1.59$, $p = .117$). Controlling for age, the results show that P300 has a significant interaction with the effect of total ETV score on individual lifetime aggression, $F(1, 61) = 7.52$, $p = .008$. While greater ETV score means greater RISQ aggression score, this effect is strongest when P300 amplitude is lowest (see Figure 1).

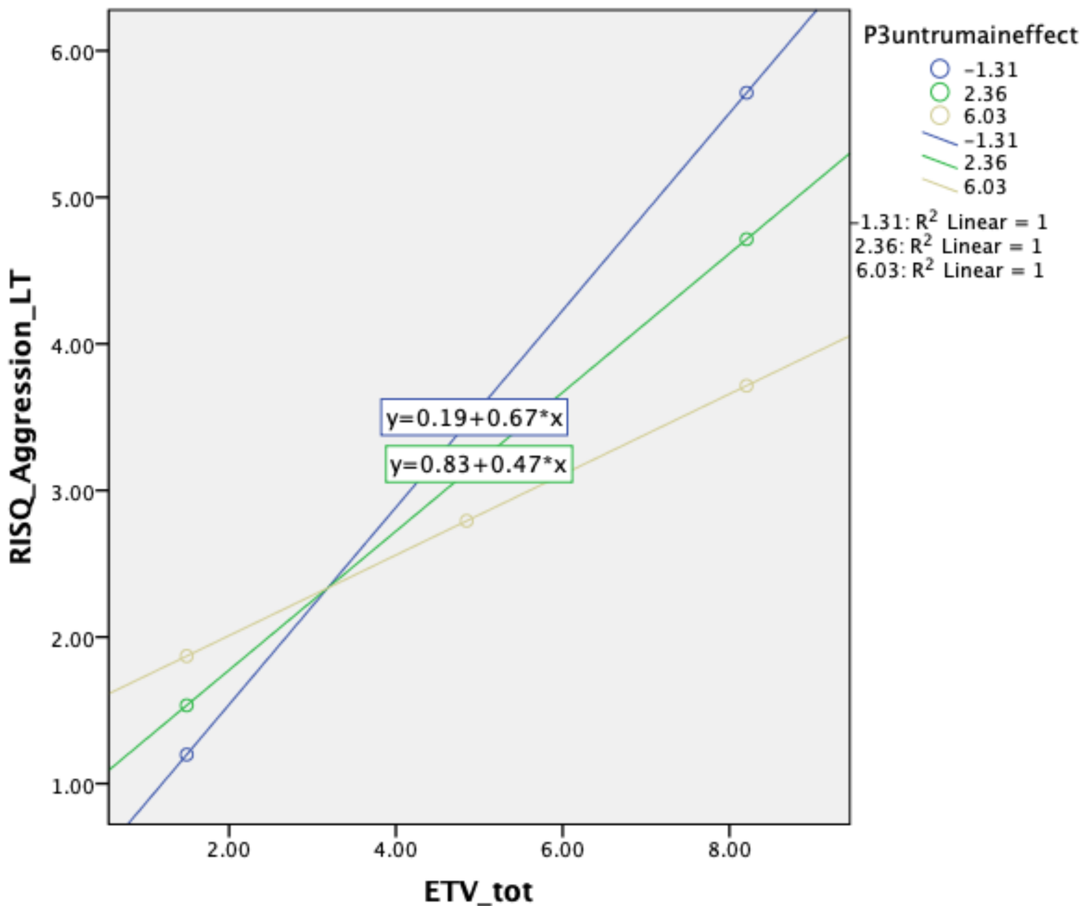

Figure 1. Higher ETV score predicts higher RISQ aggression score, or counts of aggressive behavior. P300 amplitude moderates that, showing that the strongest positive effect of ETV on aggression occurs when P300 amplitude is lowest. Individuals with higher P300 amplitude, or more neural reactivity, show a smaller effect of ETV on RISQ aggression.

Discussion

Exposure to violence has been shown to have long term effects on the social and neural functioning of individuals (Delaney-Black et al., 2002; Guerra, Rowell Huesmann, & Spindler, 2003). Trustworthiness perception has been proposed as a specific mechanism that affects social decision-making and behavior (Bonnefon, Hopensitz, & De Nys, 2013; Willis & Todorov, 2006) and may be affected by exposure to violence. The present study adds to prior research by demonstrating that ETV has a significant effect on the neural activity of individuals during both explicit and implicit trustworthiness perception. Although ETV was previously shown to correlate with aggressive behavior (Stouthamer-Loeber, Loeber, Wei, Farrington & Wikström, 2002), the present study showed that P300, as a measure of neural activity, mediates this effect, pointing to a possible pathway from adverse experiences to decision-making.

Task Performance

The results showed no significant effect of ETV on task performance, both in the explicit and implicit trustworthiness perception scenarios. In the explicit facial trustworthiness ratings, we see that people who have been exposed to violence are able to make accurate trustworthiness ratings when presented with a novel face. This is intuitive because those exposed to violence from childhood onward would likely adapt to be aware of the trustworthiness of others, especially unfamiliar faces.

The non-significant effect of ETV on implicit trustworthiness rating also makes sense when considering the environment of these individuals. Trustworthiness needs to be accurately and quickly assessed to survive in a neighborhood with high levels of violence. Therefore, the effect of exposure to violence on behavior is not due to inaccurate assessments of trustworthiness, but rather modifies something later in the decision-making process than just simple perception.

Neural Reactivity

ETV significantly increased the ERP amplitude of trustworthiness perception in both measures across the explicit condition (LPP) and implicit condition (P300). This significant neural difference demonstrates that individuals with higher exposure to violence need to allocate more neural resources to accurately rate trustworthiness, whether they are consciously doing so or not. This allocation of resources offers insight into how ETV's effect on trustworthiness perception eventually leads to more aggressive decision making and action.

A possible explanation to the results lies in previous research on ETV, which has focused on learning as an important part of connecting ETV and violent behavior. Mainly, individuals who regularly see violence learn to be desensitized to it. Once they are desensitized, they are theorized to reproduce the violent acts they have seen, as well as generalizing violence into new acts, as violence, both experienced and enacted, no longer elicits a negative response (Bandura, 1978; Catalano & Hawkins, 1996; Guerra, Huesmann, & Spindler, 2003). This adaptive response (desensitization) could allow individuals to minimize the negative effects of ETV, but removes the aversion to violence that reinforces pro-social behavior. Additionally, ETV is empirically shown to correlate with hypersensitivity to some social cues, emphasizing the differential learning that occurs (Gaylord-Harden et al., 2017).

The results fit this learning model of ETV, as the significantly increased motivated affective elaboration (LPP and P300) seen across all face types in both the explicit and implicit tasks points to hypersensitivity around trustworthiness perception. This would be adaptive for individuals who have learned to attend to cues around violence, as they experience it regularly in their lives. It would be most beneficial to accurately perceive trustworthiness when your environment is dangerous. Connecting this ETV-based significant increase in neural resources with aggressive behavior also fits the learning theory.

Looking at the moderation model, we see that the lowest P300 amplitude results in the highest levels of aggression once a certain threshold of ETV is met. It seems as though individuals on average who have experienced many counts of violence automatically allocate more resources to evaluating trustworthiness. However, individuals with high ETV who do not make that automatic extra allocation seem the most likely to react over-aggressively in situations. This is evidence that the automatic allocation of more neural resources is necessary for individuals with high ETV to make accurate trustworthiness perception judgments and then act pro-socially according to those judgments. The intersection of low P300 amplitude and high ETV represents the negative behavioral results occurring when that adjustment is not made. This breakdown of trustworthiness perception can explain the correlation between ETV and the resulting aggressive behavior. It seems that individuals who have been exposed to violence need to have a greater neural reaction to trustworthiness perception, especially when implicit, otherwise they may make incorrect ratings and act aggressively.

Over-reactivity to this social process is possibly adaptive for individuals with ETV, and those who are unable to adapt suffer the greatest consequences in pro-social behavior. Learning how to interact in an environment high in violence leads to desensitization to committing violent acts (Bandura, 1978; Catalano & Hawkins, 1996; Guerra, Huesmann, & Spindler, 2003), but should also lead to hypersensitivity to social cues. It is possible that when hypersensitivity to social cues does not occur implicitly, individuals with high ETV are more likely to misread social cues and engage in aggressive behavior. This gives a clue into the connection between the experience of violence and the enacting of it.

Limitations

There are several limitations of the study, arising from both the sample and the methodology. Firstly, the sample contained only men because of the set validated on consensus trustworthiness ratings contained only male faces and females have been shown to evaluate trustworthiness differently from men (Buchan, Croson, & Solnick, 2008; Carré, Baird-Rowe, & Hariri, 2014). The sample was not limited to white men, however, so further research should focus on in-group and out-group trustworthiness perception, and isolate race to ensure that there is no mediating effect on ratings.

A methodological limitation arose from the limit of distinct faces offered in stimuli set, as it was possible for participants to view the same face twice in both the explicit ratings task and then again in the implicit approach-avoidance task. Although there is research that implies that men do not change ratings even after repeated exposure to the same faces (Wincenciak, Dzhelyova, Perrett, & Barraclough, 2013), it is possible that familiarity could have affected the neural results.

Conclusion

This study provides preliminary evidence that exposure to violence alters the way that individuals perceive trustworthiness of novel faces, resulting in greater allocation of resources necessary to make both explicit and implicit ratings. The moderation of ERP shows the importance of examining which factors connect experience to neural mechanism, and then mechanism to real world behavior. Understanding the neural mechanism is the first step to mitigating the negative behavior, and possibly improving the social functioning of individuals exposed to violence. There has been promising research in deficit-specific interventions, which use information about neural arrears to rehabilitate behavior (Kemp & Baskin-Sommers, 2019; Javaras, Williams, & Baskin-Sommers, 2018; Brazil, van Dongen, Maes, Mars & Baskin-Sommers, 2018). As 70% of youth in the US either witness, learn about, or suffer violent acts (Finkelhor, Turner, Shattuck, & Hamby, 2013), studying this mechanism could have huge implications for therapy. ETV is correlated with increased risk of criminal justice involvement (Hawkins et al., 2000), so mechanism-specific interventions could improve rehabilitation in incarcerated populations. Overall, ETV is associated with an increased risk of violence, particularly in communities already characterized by disadvantage (Aisenberg & Herrenkohl, 2008; Finkelhor, 2009). Further study of how ETV affects trustworthiness perception and other forms of social cue reading could improve the lives of not only millions of individuals, but also help rebuild communities.

Author Contributions

Katherine Hong and Dr. Baskin-Sommers conceived and designed the analysis. Katherine Hong, Ariel Chang and RAs in the MoD lab collected the data. Katherine Hong and Dr. Baskin-Sommers performed data analysis. Katherine Hong wrote the paper.

Acknowledgments

Thank you to Dr. Arielle Baskin-Sommers for unwavering support and much needed wisdom about this thesis and beyond. This project would not have been possible without the hard work of the MoD Squad, a community that I never expected to find at Yale but one that has indelibly shaped my academic experience. Finally, thank you to Mark Sheskin and the Cognitive Science seniors who constantly remind me how beautiful and fascinating the mind can be.

References

Aisenberg, E., & Herrenkohl, T. (2008). Community Violence in Context: Risk and Resilience in Children and Families. *Journal of Interpersonal Violence, 23*, 296-315. doi:10.1177/0886260507312287

Anderson, E. (1994). The Code of the Streets. *The Atlantic Monthly*, 81-94.

Bailey, B., Delaney-Black, V., Hannigan, J., Ager, J., Sokol, R., & Covington, C. (2005). Somatic Complaints in Children and Community Violence Exposure. *Journal Of Developmental & Behavioral Pediatrics*, *26*(5), 341-348. doi: 10.1097/00004703-200510000-00001

Bandura, A. (1978). Social learning theory of aggression. *Journal of communication, 28*(3), 12-29. doi:10.1111/j.1460-2466.1978.tb01621.x

Baskin, D., & Sommers, I. (2014). Trajectories of Exposure to Community Violence and Mental Health Symptoms Among Serious Adolescent Offenders. *Criminal Justice And Behavior*, *42*(6), 587-609. doi: 10.1177/0093854814556882

Bonnefon, J., Hopensitz, A., & De Nys, W. (2013). The Modular Nature of Trustworthiness Detection. Journal of Experimental Psychology: General, 142-150. doi:http://dx.doi.org/10.1037/a0028930

Brazil, I., van Dongen, J., Maes, J., Mars, R., & Baskin-Sommers, A. (2018). Classification and treatment of antisocial individuals: From behavior to biocognition. *Neuroscience & Biobehavioral Reviews*, *91*, 259-277. doi: 10.1016/j.neubiorev.2016.10.010

Buchan, N. R., Croson, R. T., & Solnick, S. (2008). Trust and gender: An examination of behavior and beliefs in the Investment Game. Journal of Economic Behavior & Organization, 68(3), 466-476.

Campbell, C., & Schwarz, D. (1997). Prevalence and impact of exposure to interpersonal violence among suburban and urban middle school students. *Pediatrics*. doi: https://www.ncbi.nlm.nih.gov/pubmed/8784363

Catalano, R. F., & Hawkins, J. D. (1996). The social development model: A theory of antisocial behavior. In J. D. Hawkins (Ed.), *Deliquency and Crime: Current Theories* (pp. 149-197). Cambridge: Cambridge University Press.

Delaney-Black, V., Covington, C., Ondersma, S., Nordstrom-Klee, B., Templin, T., & Ager, J. et al. (2002). Violence Exposure, Trauma, and IQ and/or Reading Deficits Among Urban Children. *Archives Of Pediatrics & Adolescent Medicine*, *156*(3), 280. doi: 10.1001/archpedi.156.3.280

Finkelhor, D. (2009). *Children's exposure to violence: A comprehensive national survey*: DIANE Publishing.

Finkelhor, D., Turner, H. A., Shattuck, A., & Hamby, S. L. (2013). Violence, crime, and abuse exposure in a national sample of children and youth: an update. *JAMA pediatrics, 167*(7), 614-621.

Foti, D., Hajcak, G., & Dien, J. (2009). Differentiating neural responses to emotional pictures: evidence from temporal-spatial PCA. Psychophysiology, 46(3), 521-530.

Fowler, P., Tompsett, C., Braciszewski, J., Jacques-Tiura, A., & Baltes, B. (2009). Community violence: A meta-analysis on the effect of exposure and mental health outcomes of children and adolescents. *Development And Psychopathology*, *21*(01), 227. doi: 10.1017/s0954579409000145

Gaylord-Harden, N. K., Bai, G. J., & Simic, D. (2017). Examining a Dual-Process Model of Desensitization and Hypersensitization to Community Violence in African American Male Adolescents. *J Trauma Stress, 30*(5), 463-471. doi:10.1002/jts.22220

Guerra, N., Rowell Huesmann, L., & Spindler, A. (2003). Community Violence Exposure, Social Cognition, and Aggression Among Urban Elementary School Children. *Child Development*, *74*(5), 1561-1576. doi: 10.1111/1467-8624.00623

Guo, X., Zheng, L., Wang, H., Zhu, L., Li, J., & Wang, Q. et al. (2013). Exposure to violence reduces empathetic responses to other's pain. *Brain And Cognition*, *82*(2), 187-191. doi: 10.1016/j.bandc.2013.04.005

Hajcak, G., Dunning, J. P., & Foti, D. (2009). Motivated and controlled attention to emotion: time-course of the late positive potential. Clinical Neurophysiology, 120(3), 505-510.

Hawkins, J. D., Herrenkohl, T. I., Farrington, D. P., Brewer, D., Catalano, R. F., Harachi, T. W., & Cothern, L. (2000). Predictors of Youth Violence. Juvenile Justice Bulletin.

Javaras, K.†, Williams, M.* & Baskin-Sommers, A.R.† (2018). Behavioral interventions for constructs relevant to conscientiousness. Personality Disorders: Theory, Research and Treatment, 10, 13-24. †These authors contributed equally to this work.

Kemp, E. & Baskin-Sommers, A. (in press). Using an experimental therapeutics approach to target psychopathy. In S. Dimidjian (Ed.). Evidence-Based Practice in Action. Guilford Press: New York, NY.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). Motivated attention: Affect, activation, and action. In P. J. Lang, R. F. Simons, & M. T. Balaban (Eds.), *Attention and orienting: Sensory and motivational processes* (pp. 97-135). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Lischke, A., Junge, M., Hamm, A., & Weymar, M. (2017). Enhanced Processing of Untrustworthiness in Natural Faces With Neutral Expressions (Vol. 18).

Marzi, T., Righi, S., Ottonello, S., Cincotta, M., & Viggiano, M. P. (2014). Trust at first sight: evidence from ERPs. Soc Cogn Affect Neurosci, 9(1), 63-72. doi:10.1093/scan/nss102

Monahan, K., King, K., Shulman, E., Cauffman, E., & Chassin, L. (2015). The effects of violence exposure on the development of impulse control and future orientation across adolescence and early adulthood: Time-specific and generalized effects in a sample of juvenile offenders. *Development And Psychopathology*, *27*(4pt1), 1267-1283. doi: 10.1017/s0954579414001394

Mrug, S., & Windle, M. (2008). Mediators of Neighborhood Influences on Externalizing Behavior in Preadolescent Children. *Journal Of Abnormal Child Psychology*, *37*(2), 265-280. doi: 10.1007/s10802-008-9274-0

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. Proceedings of the National Academy of Sciences, 105(32), 11087-11092.

Sadeh, N., & Baskin-Sommers, A. (2016). Risky, Impulsive, and Self-Destructive Behavior Questionnaire (RISQ): A Validation Study. *Assessment*, *24*(8), 1080-1094. doi: 10.1177/1073191116640356

Schubiner, H., Scott, R., & Tzelepis, A. (1993). Exposure to violence among inner-city youth. *Journal Of Adolescent Health*, *14*(3), 214-219. doi: 10.1016/1054-139x(93)90008-d

Selner-O'Hagan, M., Kindlon, D., Buka, S., Raudenbush, S., and Earls, F. (1998). Assessing exposure to violence in urban youth. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 39(2), 215-224.

Stouthamer-Loeber, M., Loeber, R., Wei, E., Farrington, D., & Wikström, P. (2002). Risk and promotive effects in the explanation of persistent serious delinquency in boys. *Journal Of*

Consulting And Clinical Psychology, *70*(1), 111-123. doi: 10.1037//0022-006x.70.1.111

Todorov, A., Baron, S. G., & Oosterhof, N. N. (2008). Evaluating face trustworthiness: a model based approach. Social Cognitive and Affective Neuroscience, 3(2), 119-127. doi:10.1093/scan/nsn009

Willis, J., & Todorov, A. (2006). First Impressions: Making Up Your Mind After a 100-Ms Exposure. Psychological Science, 17(7), 592-598.

Wincenciak, J., Dzhelyova, M., Perrett, D. I., & Barraclough, N. E. (2013). Adaptation to facial trustworthiness is different in female and male observers(). Vision Research, 87, 30-34. doi:10.1016/j.visres.2013.05.007

Yang, D., Qi, S., Ding, C., & Song, Y. (2011). An ERP study on the time course of facial trustworthiness appraisal. Neurosci Lett, 496(3), 147-151. doi:10.1016/j.neulet.2011.03.066

Zachary, R. A. (1986). Shipley Institute of Living Scale: Revised Manual. Western Psychological Services, Los Angeles, CA.