Exposure to violence is not associated with accuracy in forecasting conflict outcomes

Abstract

Exposure to harsh or unpredictable environments can impair social and cognitive functioning. However, people may also develop enhanced abilities for solving challenges relevant in those environments ('hidden talents'). In the current study, we explored the associations between people's ability to accurately forecast conflict outcomes and their past and current experiences with violence. To do so, we used dynamic, real-world videos with known outcomes, rather than static, artificial stimuli (e.g., vignettes) with unknown outcomes, as previous research has done. We conducted a preregistered study in the Netherlands that included a final sample of 127 participants: 63 from a community sample and 64 college students. We found no support for our core hypothesis that people who experienced more violence are more accurate in forecasting conflict outcomes. Thus, we did not find support for hidden talents, contributing to an evidence base that was already mixed and inconclusive. We did find support for our auxiliary hypothesis that college students would wear 'rose-colored glasses', underestimating the number of conflicts that would escalate into fights. Contrary to our other two auxiliary hypotheses, the community sample did not overestimate the number of conflicts that would escalate into fights, and people who have experienced more violence were not more likely to predict that conflicts will escalate into fights. These findings have implications for the literature on hostile attribution bias, which shows that people with more exposure to violence more likely interpret the ambiguous actions of others as hostile. Whereas this pattern is often attributed to negativity bias in people with more exposure to violence, it might also reflect rose-colored glasses on people living safer lives.

Keywords: development, adversity, violence, harsh parenting, predicting conflict outcomes

Exposure to violence is not associated with accuracy in forecasting conflict outcomes

"Don't push me 'cause I'm close to the edge, I'm trying not to lose my head" (Grandmaster Flash and the Furious Five, 1982)

People who grow up in harsh or unpredictable environments are more likely to develop impairments in their social and cognitive abilities than people who grow up in more safe and stable environments (Duncan et al., 2017; Evans et al., 2013). These impairments, in turn, are associated with important outcomes later in life, such as worse health, wealth, and educational outcomes (reviewed in Ellis et al., 2017). Knowledge about impairments is extremely valuable and has generated policies and interventions to prevent and repair deficits, some of which have improved the lives of many people (e.g., Leve et al., 2012). However, we have argued that any approach that focuses exclusively on impairments is incomplete, because it overlooks abilities that are enhanced by adversity through adaptive developmental processes ('hidden talents').

The hidden talents approach

The hidden talents approach examines whether people exposed to adversity do not only develop deficits, but enhanced abilities as well (Ellis et al., 2017; Frankenhuis & de Weerth, 2013; Frankenhuis, Young, et al., 2020). It proposes that people in harsh and unpredictable environments may develop enhanced skills and abilities for solving challenges relevant in those environments. These abilities are called 'hidden talents', because they tend to remain outside the purview of conventional testing. For example, the tendency to rapidly shift attention might be useful in unpredictable environments, because it allows people to anticipate and act on sudden threats or opportunities. However, this tendency may also reduce sustained attention, which is

beneficial in formal educational settings (Mittal et al., 2015). Studies of hidden talents have explored a range of social and cognitive abilities, including danger detection (Pollak, 2008), memory (Belsky et al., 1996; Goodman et al., 2009), empathic accuracy (Bjornsdottir et al., 2017; Kraus et al., 2012), and executive functioning (Fields et al., 2021; Mittal et al., 2015; Nweze et al., 2020; Young et al., 2018; Vandenbroucke et al., 2016). However, the evidence base for each hidden talent is small or mixed and inconclusive (Frankenhuis, Young, et al., 2020).

In this article, we explore the associations between previous experience and accuracy in perceptions of social interactions. Prior research has focused on social information processing abilities of people from less supportive environments (Crick & Dodge, 1994). For example, some studies suggest that people growing up in low socioeconomic conditions may show enhanced empathic accuracy (Bjornsdottir et al., 2017; Kraus et al., 2012). Other work found that, on some measures but not others, people who have experienced greater exposure to violence show intact or enhanced memory for social-dominance relationships (Frankenhuis, de Vries, et al., 2020). In addition, studies in children that were severely neglected and/or abused have demonstrated greater speed in recognizing fearful facial expressions and accuracy in recognizing angry facial expressions (Masten et al., 2008; Pollak et al., 2009). However, this evidence is mixed as well. For instance, in a socioeconomically diverse Dutch sample, we did not find evidence that people exposed to more violence detected angry (and sad) faces more accurately, and people who had such experiences actually responded more slowly to angry faces (Frankenhuis & Bijlstra, 2018). An open question is whether these mixed findings can be attributed to transparent and complete reporting (which should lead to mixed findings, even when there is a true effect; Francis, 2014;

Lakens & Etz, 2017; Schimmack, 2012) or due to methodological issues, such as differences between study populations, measurement challenges, and limited statistical power.

The current study

The current study extends the existing literature by examining whether people who have been exposed to more violent environments are better able to 'forecast' whether dyadic conflicts will escalate into fights. As opposed to previous studies that have typically used static, artificial, ambiguous stimuli–such as vignettes, pictures of neutral faces, and acted scenarios–the current study used dynamic, real-world, unambiguous videos of confrontations. Some research suggests that movie-watching, especially of videos with social content, outperforms other measures in predicting cognitive and emotional traits (Finn & Bandettini, 2021). In addition, whereas people in social information processing studies are typically asked to interpret scenarios and predict responses without there being a correct answer, the current study examines accuracy in predicting known outcomes (to the experimenter).

We hypothesized that in environments in which violence is relatively more common and consequential, it is particularly important for people to be accurate at predicting whether a social conflict will escalate into a fight. To our knowledge, no prior research has tested this hypothesis. However, the idea that people in dangerous environment develop enhanced abilities for 'reading dangerous situations' appears in qualitative ethnographies by anthropologists, economists, and people who have themselves lived experience in violent contexts (Keiser, 1979; Shakur, 2007; Venkatesh, 2008). In his book *The vice lords: Warriors of the streets*, Keiser (1979) notes:

> Individuals often argue in the course of social interaction, and these arguments do not always signal enmity. (…) it is not always clear whether an argument is 'consequential' or not, and there are (…) subtle cues which (…) indicate if an identity of enmity is being

assumed. (…) Possibly, they consist of such things as facial expressions and tones in the

voice. (…) Being adept at responding properly is one of the things that constitutes

'knowing what's happening,' or (…) 'knowing how to live on the streets'. (p. 45)

Our study focuses on three types of exposures to violence in relation to people's ability to

forecast conflict outcomes: (a) past and current (passive) exposure to neighborhood violence; (b)

past and current (active) involvement in violence; and (c) harsh parenting during childhood and

adolescence. To obtain variation in these exposures, we sampled participants from two different

groups. One group of community participants who lived in relatively challenging conditions for

Dutch standards, in that people were more likely to need governmental support in order to meet

their basic needs (e.g., food, housing, safety). The community sample is diverse. Nonetheless, we

anticipated that on average members of the community sample had experienced higher levels of

adversity, including higher levels of exposure to violence in their past or current environments

(and thus termed this sample the 'high-risk[1] group'). The other group was a college student

sample, who we expected to have experienced lower levels of violence (the 'low-risk group').

Predictions

To test our core hypothesis, we examined whether the high-risk group and individuals

who have experienced more violence (i.e., exposure to and/or involvement in violence) in their

past or current environments are more accurate in forecasting conflict outcomes (i.e., whether

conflicts escalated into fights) [*Hypothesis 1*]. Although we expected people who experienced

more violence to be more accurate in forecasting conflict outcomes, we included the general

---

[1] In our preregistration, we used the term 'high-risk group' to refer to our socioeconomically diverse community sample, and 'low-risk group' to refer to college students. This terminology is commonly used in the field. However, we have moved away from it in our recent work, because it may be stigmatizing for community participants. Nonetheless, to ensure a match between the preregistration and the paper, we will use this terminology when stating our predictions.

impairment view in our hypotheses (which predicts the opposite pattern) to ensure our analyses would also be informative about deficit models, which are predominant in the literature.

We also had auxiliary hypotheses. In terms of absolute accuracy, we predicted that the high-risk group would have a 'negativity bias' (i.e., overestimate the number of conflicts that would escalate into fights; Cacioppo et al., 1999; Castro & van Dijk, 2017; De Castro et al., 2002; Rozin & Royzman, 2001) [*Hypothesis 2*], whereas the low-risk group would wear 'rose-colored glasses' (i.e., underestimate the number of conflicts that would escalate into fights; Laible et al., 2014; Nelson & Crick, 1999) [*Hypothesis 3*]. In terms of relative outlook, we predicted that the high-risk group and individuals who have experienced more violence in their past or current environment would forecast more conflicts to escalate into physical fights [*Hypothesis 4*].

Methods

We preregistered the hypotheses, sample size, and statistical analyses for this study at the Open Science Framework: https://osf.io/uijwe/?view_only=aea332f69f8d49d0a56a8f86e5c4f9c9 (under Fight Study). This preregistration is less clear and detailed than current best practices. The reason is that we developed this preregistration in 2014 and uploaded it on 2015-01-27, when we had less expertise than we do now, in part because there were fewer templates and examples. In what follows, we will call attention to aspects of our study that should have been clearer in our preregistration as well as deviations from our preregistration.

Our materials, data, and analysis script are available at: https://osf.io/2rt95/?view_only=006fc220639e4e16bbb4a2e93f824c87. We do not have copyright permission to upload the videos we used under a permalink. Therefore, we can only

share these videos upon request. This study was approved by the Ethics Committee of the

Faculty of Social Sciences, Radboud University (CSW2014-1310-250).

Participants

First, we conducted a pilot study of 45 participants in the United States to evaluate and

improve the study design (e.g., videos) and estimate effect sizes to determine our target sample

size (for recent criticism of this approach; Albers & Lakens, 2018). This pilot sample consisted

of 21 community participants tested in an employment development center (7 females, M_{age} =

20.33, SD = 2.74, range = 16-28) and 24 college students tested in the lab (16 females, M_{age} =

19.58, SD = 1.32, range = 18-23).

For the Dutch sample we initially tested 158 participants. This original sample included

66 high-risk and 92 low-risk participants. Of the high-risk group, we excluded three participants:

one whose behavior suggested that s/he was under influence of a mind-altering substance, one

who indicated not understanding the first three trials of the fight prediction task, and one who

indicated not having read the task instructions. As for the low-risk group, we excluded the 28

participants tested most recently–based only on their date and time of participation, without

having seen their data–because we accidentally tested them beyond our preregistered sample

size. To not waste data, we ran our main analyses including these 28 participants as a post-hoc

sensitivity analysis (i.e., N = 155). The results of this analysis were qualitatively similar to those

with our preregistered sample size reported below.

Our final sample thus included 127 participants: 63 participants in the high-risk

community sample (22 males, M_{age} = 40.11, SD = 13.44, range = 18-65) and 64 participants in

the low-risk student sample (26 males, M_{age} = 22.83, SD = 5.55, range = 18-61). Therefore, we

had one (community) participant less than our preregistered sample size of 128, that was based

on a power analysis for independent sample t-tests in G*Power (Erdfielder et al., 1996) with a medium effect size ($d = 0.5$), 5% alpha-level, and 80% statistical power. The community sample was recruited through several community organizations that provide support for people living in relatively challenging conditions for Dutch standards. Most of our participants were exposed to stressors or adversity at the time of testing, and/or had been exposed to adverse events in the past (e.g., job and/or housing insecurity, exposure to violence). Participants were recruited through flyers and personal communication.

Materials

Fight prediction. We developed a video task to measure fight forecasting accuracy[2]. Participants watched videos of real-life conflicts and were asked to judge whether each of these conflicts would lead to physical fights. We obtained 16 videos from YouTube showing dyadic confrontations between men (from diverse ethnic backgrounds), using six selection criteria: (a) the video showed a tense situation that may result in a fight; (b) the video was shorter than 30 seconds (or could be shortened to 30 seconds or less); (c) the video depicted male-versus-male arguments or fights; (d) the video does not show a group fight (i.e., we allowed for bystanders, but not for more than two people being involved as aggressors); (e) the video is of reasonable quality (e.g., not too dark or grainy to see what is happening); (f) the video had not received thousands of views (to reduce the chance that participants would be familiar with the videos).

In eight of the final 16 videos, conflicts culminated into fights, and in the other eight they did not, with a caveat. For the videos we deemed non-fights, we do not know with certainty that no fight occurred after the recording stopped. However, we thought this highly unlikely, because

---

[2] Our preregistration did not state explicitly that we would use a video paradigm. It only stated: "Participants predict whether dyadic conflicts will escalate into fights." We should have made this explicit. This video task was our only measure of the ability to forecast conflict outcomes.

in most of these videos, the men clearly parted ways; for instance, one man departed the subway while the other continued traveling, or one biked off into the distance while the other remained in the location of the conflict. This uncertainty about video endings is a consequence of using ecologically valid, rather than experimentally controlled, stimuli (e.g., acted scenarios). Participants were not told about this 50-50 split in fight and no-fight videos. Participants viewed the videos sequentially and in the same order. Each video would start by showing a conflict and stop just before it became clear whether this conflict would escalate into a fight. The duration of the videos was 10-20 seconds.

Before the task, participants were told that they would watch 16 short videos that each showed a real conflict between two people. They were also told that some of these conflicts had led to a fight while others did not and that they would see a part of the video, but not the end of it. They were then asked to predict as well as they could whether the conflict did or did not lead to a fight after the video ended. Participants answered two questions after watching each video: (1) "Will they fight?" (yes/no), and (2) "Have you ever seen this video before?" (yes/no). To ensure that all participants interpreted the first question in the way we intended and in the same way, we provided a definition of fighting. This definition was displayed while participants rated each of the videos: "a violent confrontation, with at least one person physically harming the other by hitting, kicking, biting, etc. (pushing alone does not count as fighting)." After participants had responded they could not watch the remainder of the video, nor were they told at any point whether or not their answers were correct, nor did we ask participants who they expected to win the fight if the conflict were to escalate (for recent research on this topic, see Aung et al., 2021; Fink et al., 2019; Kupor et al., 2019; Lane & Briffa, 2020).

Neighborhood violence. We used the Neighborhood Violence Scale (for a description of the development of this scale, see Frankenhuis et al., 2018) to measure childhood exposure to neighborhood violence (e.g., 'Crime was common in the neighborhood where I grew up') and current exposure to neighborhood violence (e.g., 'In my neighborhood, most people feel unsafe walking alone after dark'). Both of these subscales consist of seven identical items, except in referring to childhood and adolescence (before the age of 18; $\alpha = 0.85$) and current experiences as an adult ($\alpha = 0.88$). We rated all items on a scale from 1 (*strongly disagree*) to 7 (*strongly agree*). We computed an average score for each subscale for each participant, where a higher score represents more neighborhood violence.

Involvement in violence. We used a four-item subset of the Youth Risk Behavior Survey (Eaton et al., 2012) to measure direct involvement in violence[3]. Here, we asked participants about the frequency of involvement in physical fights that necessitated treatment for injuries as a minor (14-17 years of age), as well as in the last year, using two items that were rated on a scale from 1 (*0 times*) to 5 (*6+ times*). Using the other two items, we also asked participants to indicate the frequency of their childhood and current involvement in a physical fight broadly (regardless of injuries), on a scale from 1 (*0 times*) to 8 (*12+ times*). We computed two composite scores for the two periods by averaging the two childhood and current items[4], in which 7 scores above five

---

[3] We preregistered which four items on the preregistration page that provides an overview of all questionnaires used in the larger test battery (p. 1) in which the current study was embedded.

[4] As both subscales contained only two items with positively skewed and leptokurtic distributed scores, we used Kendall's tau instead of Chronbach's alpha to compute correlations between the childhood ($r_\tau = .39$, $p < .001$) and current items ($r_\tau = .25$, $p = .004$). These coefficients might be lower than desired. However, components of composite scores do not necessarily have to be highly correlated, given the rarity of the events described in these items (here the involvement in physical fights in a broad sense and in those that necessitated treatment) and the common practice to 'sum' such–correlated or uncorrelated–events to a compositive score.

on 8-point scales were truncated to a value of exactly five (i.e., a score of 5 represents 6+ times

on both scales).

Harsh parenting. We used the harsh-coercive parenting scale of the Parenting

Questionnaire (Ellis et al., 2012) to measure harsh parenting. This scale included 8 items (4 for

each parent, e.g., 'My father/mother swore (cursed) at me'; $\alpha = 0.83$). We asked participants to

indicate on a 4-point scale the degree to which an item described each of their parents in the first

16 years of their life (*0 = very unlike, 3 = very like*). We computed a single average score that

combined the scores for both parents for each participant, where a higher score indicates harsher

parenting. If participants had one parent only, we computed the average score based on the items

concerning this parent.

Exploratory instruments. We measured five constructs for exploratory analyses,

namely age (in years), gender (female, male, or other), childhood and current material needs, and

perceived life expectancy. We used a version of the material needs scale developed by Conger et

al. (1994) to measure childhood and current material needs (i.e., adequacy of resources to make

basic ends meet). This scale included eight items, with four for childhood and adolescence (until

the age of 16; $\alpha = 0.96$), and four identical items for current experiences as an adult (e.g., 'You

have enough money to afford the house you need'; $\alpha = 0.91$). All were rated on a 7-point scale (*0

= strongly disagree, 7 = strongly agree*). We measured perceived life expectancy using a scale

that we developed in the majority of the community sample and about half of the student sample,

because we added this scale later. This scale consisted of seven items (e.g., 'Do you think you

will live to be 60 or older?'; $\alpha = 0.72$). These were rated on 5- or 7-point scale, with a higher

score indicating higher perceived life expectancy. We analyzed these five constructs using

identical statistical procedures as we used in our core and auxiliary analyses.

Procedure

We have collected data from March 17th, 2014, through November 13th, 2015. Before

participation, we obtained informed consent from all participants. We tested the student sample

in a test cubicle at the university on a 24-inch desktop, and the community sample in a separate

room in their community organization building on a 17-inch laptop. All stimuli were displayed in

the center of the screen. After we provided verbal instructions, participants completed all of the

questionnaires before starting with the fight task. Participants completed the questionnaires and

task by themselves, while having the possibility to ask the experimenter clarification questions.

All participants in both samples received a financial compensation of 10 or 15 euro at the end of

the session, depending on the length of the session (60 or 90 minutes, respectively). The duration

of the session varied as a function of participants' pace of reading and the number of questions

they had about the instructions or task.

Participants took 10 to 15 minutes to complete the fight task. As noted, however, the

entire session lasted 60 to 90 minutes. This is because the fight task was part of a test battery that

included three other, non-related tasks (described in Frankenhuis & Bijlstra, 2018; Frankenhuis,

de Vries, et al., 2020; Frankenhuis et al., 2018). As participants always completed the fight task

before completing these other three tasks, it is not possible for those other tasks to have affected

performance on the fight task (e.g., due to fatigue or changes in mood).

Results

Based on Shapiro-Wilk tests of normality, we rejected the hypothesis that some of our

variables for confirmatory analyses were drawn from a normal distribution. Therefore, we have

reported medians in addition to means and conducted non-parametric analyses such as Kendall's

tau for correlations. For all hypothesis tests, we have used $p < .05$ as a criterion for significance.

We have used two-tailed tests to be able to examine both enhanced and impaired performance.

False discovery rate control

For Hypotheses 1 and 4, we tested six predictions: one comparing the student and the

community sample and five for the continuous predictors (two measures of exposure to violence,

two of involvement in violence, and one of harsh parenting). To control for multiple testing of

Hypotheses 1 and 4, we applied the sequential Holm–Bonferroni correction (Holm, 1979). We

applied this correction per hypothesis, as recommended by Lakens (2016), for these six tests. For

Hypotheses 2 and 3, we tested one prediction, thus we did not control for multiple comparisons.

Signal Detection Theory

We used signal detection theory to analyze accuracy and bias in participants' responses

(Green & Swets, 1974; Macmillan & Creelman, 2005). Participants may rate non-fights as non-

fights (correct rejections), non-fights as fights (false alarms), fights as non-fights (misses), and

fights as fights (hits). The proportion of hits and false alarms were used to compute accuracy d'

and bias c (see Stanislaw & Todorov, 1999, for formulas). It is generally useful to distinguish

between accuracy and bias, and absolutely necessary when there is a different number of signal

(fight) and noise (no-fight) trials for some or all participants. In the extreme, if 95% of all trials

depict conflicts that will escalate into fights, people with low thresholds for predicting a conflict

to escalate are likely to respond correctly, when guessing at random. In such cases, a measure of

'proportion correct' confounds accuracy and bias. Signal detection theory resolves this issue by

computing accuracy independently of bias, taking into account the number of each type of trial.

The parameter d' represents accuracy in discriminating between signal (fight) and no-

signal (no-fight) trials, where a higher d' indicates greater accuracy and has a lower bound of

zero[5]. Criterion *c* represents the threshold for recognizing a signal (i.e., fight) trial: a *c* of zero indicates no bias, a negative *c* a low threshold (fight bias; consistent with hostile attribution bias; De Castro & van Dijk, 2017; De Castro et al., 2002), and a positive *c* a high threshold (no-fight bias; consistent with rose-colored glasses; Laible et al., 2014; Nelson & Crick, 1999). As some participants attained extreme scores (i.e., had 0% or 100% hits and/or false alarms), we applied the log-linear method to all participants' values before computing *d'* and *c* to improve estimates (Brown & White, 2005; Hautus & Lee, 2006; Snodgrass & Corwin, 1988; Stanislaw & Todorov, 1999). We conducted signal detection analyses using R version 4.0.2 (R Core Team, 2020).

During data collection–thus after having preregistered our study–we learned that it would be informative to analyze the data using Generalized Linear Mixed Models, because such models include random effects in addition to fixed effects, and hence reduce type 1 and 2 error rates (Barr et al., 2013; Watkins & Martire, 2015). Therefore, we report these non-preregistered analyses in the Appendix, Section 1. These results showed the same pattern as those of our preregistered analyses described below, with one exception: people with higher levels of past involvement in violence predicted that more conflicts would escalate into fights than people with lower levels of past involvement in violence. This result provides some support for our auxiliary hypothesis that people who have experienced more violence in their past or current environment would forecast more conflicts to escalate into fights. However, we believe that this one exception should not substantially change our qualitative conclusion about this hypothesis, because the other predictions of this hypothesis were not supported (see below).

---

[5] Our preregistration did not state explicitly that we would use *d'* to represent accuracy. At the time of our preregistration, we might have also considered other options, such as the proportion of correct predictions. However, before analyzing the data, we realized that 'proportion correct' would confound accuracy and bias (e.g., if there is an unequal number of signal and noise trials due to missing data), and therefore we decided to use *d'*.

Bayes Factors

We report Bayes Factors (BF) as well as *p*-values (Jeffreys, 1961; Rouder et al., 2009). We computed BFs using JASP (JASP Team, 2020). *BF*s describe the likelihood ratio of the data, typically given a null hypothesis (H_0) and an alternative hypothesis (H_1). For example, if $BF_{10} = 5$ (or $BF_{01} = 0.20$), the data are five times more likely to have occurred under H_1 than under H_0. As there are no Bayesian Welch's *t*-tests available in JASP, we report student *BF*s for these tests instead.

Preliminary analyses

United States Sample

We collected the United States sample data as a (non-preregistered) pilot study. The goal of this pilot study was to refine our study design (instructions, exclusion criteria), not to conduct inferential statistics. Accordingly, we provide only descriptive statistics for this sample (Table 1). The average *d'* in the student sample ($M = 0.18$, $SE = 0.09$, $Mdn = 0.28$) was slightly lower than the average *d'* in the community sample ($M = 0.28$, $SE = 0.13$, $Mdn = 0.28$). This pattern raises the possibility, which we will test later in the Dutch sample, that the community sample is slightly more accurate in discriminating between fight and no-fight trials. The average *c* in the community sample ($M = -0.22$, $SE = 0.11$, $Mdn \approx 0.00$) was much lower than the average *c* in the student sample ($M = 0.17$, $SE = 0.07$, $Mdn = 0.14$). This pattern raises the possibility, which we will test later in the Dutch sample, that the community sample had a fight bias, whereas the student sample had a no-fight bias.

Table 1

Descriptive Statistics for the United States Pilot Sample

Variable	Community (n = 21)			Students (n = 24)		
	Mean (SD)	Median	Range	Mean (SD)	Median	Range
Childhood neighborhood violence	3.93 (2.03)	4.00	1.00-7.00	1.85 (0.88)	1.57	1.00-4.71
Current neighborhood violence	3.27 (1.61)	3.57	1.00-5.86	2.89 (1.50)	2.50	1.29-6.00
Childhood involvement in violence	1.83 (0.97)	1.00	1.00-3.50	1.25 (0.66)	1.00	1.00-4.00
Current involvement in violence	1.52 (0.66)	1.00	1.00-3.00	1.04 (0.14)[a]	1.00[a]	1.00-1.50[a]
Harsh parenting	1.84 (0.76)	1.63	1.00-4.00	1.36 (0.43)	1.25	1.00-2.50

[a] For current involvement in violence, there was one data point missing in the student sample.

Dutch Sample

The descriptive statistics of the confirmatory Dutch study can be found in Table 2. As noted, this study was part of a test battery, and so we have reported similar descriptive statistics for part of the current sample in previous papers (Frankenhuis & Bijlstra, 2018; Frankenhuis et al., 2018; Frankenhuis, de Vries, et al., 2020). We expected moderate or high positive correlations between our independent variables (in the range 0.4–0.8): harsh parenting, childhood and current passive exposure to neighborhood violence, and childhood and current active involvement in violence). Therefore, we decided a priori (i.e., before conducting any analyses) to

analyze these variables in distinct models. Seven out of 10 predictor pairs were indeed

significantly positively correlated, with Kendall's tau ranging from .215 to .403 (all p-values $\leq$

.002, BFs_{+0} ranging from 132 to >1000000), which is lower than we expected. Also, three

correlations were not statistically significant, specifically those concerning current involvement

in violence. This variable had a limited range and was not significantly correlated with childhood

neighborhood violence ($r_\tau = .08$, $p = .266$, $BF_{01} = 3.42$), current neighborhood violence ($r_\tau = .14$,

$p = .056$, $BF_{01} = 0.59$), or harsh parenting ($r_\tau = .06$, $p = .436$, $BF_{01} = 5.25$).

Table 2
Descriptive Statistics for the Dutch Sample

	Community (n = 63)			Students (n = 64)			Difference between samples		
Variable	Mean (SD)	Median	Range	Mean (SD)	Median	Range	W	p-value	1-sided BF$_{+0}$
Childhood neighborhood violence	3.07 (1.20)	3.00	1.00-6.43	2.26 (1.06)	2.00	1.14-6.00	4.03	<.001	453.81
Current neighborhood violence	3.89 (1.20)	3.86	1.14-6.14	2.58 (1.00)	2.43	1.00-5.71	6.69	<.001	>10000
Childhood involvement in violence	1.59 (0.82)	1.00	1.00-4.00	1.20 (0.34)	1.00	1.00-2.50	3.51	<.001	89.80
Current involvement in violence	1.10 (0.25)	1.00	1.00-2.00	1.05 (0.19)	1.00	1.00-2.00	1.21	.228	0.65
Harsh parenting [a]	1.77 (0.86)	1.50	1.00-5.00	1.41 (0.41)	1.25	1.00-2.75	2.92	.004	20.63

Note. We report one-sided BFs when comparing the adversity exposures of the student and community samples, because a priori we expected the community sample to have experienced higher levels of adversity. This prediction is the same for deficit models and adaptation theories.

[a] For harsh parenting, there were four missing values in the community sample.

Preliminary analyses: Accuracy *d'*

A Wilcoxon signed rank test indicated that the average *d'* ($M = .21$, $SE = .05$) was significantly higher than zero ($W = 3307.50$, $p < .001$, $BF_{10} = 123.65$). This indicates that on average for the two samples combined, participants did better than chance. Computing this test separately for each group indicated that *d'* did not differ significantly from zero in the community sample ($W = 802.00$, $p = .060$, $BF_{10} = 1.59$), but was significantly higher than zero in the student sample ($W = 861.00$, $p = .002$, $BF_{10} = 51.36$). Table 3 shows the number of hits, false alarms, misses, and correct rejections used to compare the student and community samples.

Table 3

Means (with Standard Deviations in Parentheses) for the Number of Hits, False Alarms, Misses, and Correct Rejections for the two Dutch Samples

Response	Community sample		Student sample	
	Fight video	No-fight video	Fight video	No-fight video
Fight	4.14 (1.80)	3.60 (1.50)	3.94 (1.67)	3.03 (1.13)
No fight	3.86 (1.80)	4.40 (1.50)	4.06 (1.67)	4.97 (1.13)

Note. The mean proportions of correct responses were .53 for the community sample and .56 for the student sample, respectively.

Core analyses: Accuracy *d'*

Welch's *t*-test did not indicate a significant difference in *d'* between the community

sample (*M* = .16, *SE* = .08, *Mdn* = .28) and the student sample (*M* = .26, *SE* = .07, *Mdn* = .28), *t*

= -.90, *p* = .370, 95% CI [-.30, .11], student BF_{01} = 3.65 [*Hypothesis 1, group-level prediction*].

We show for the distributions of these data using box plots in the Appendix, Section 2.

Combining both samples, there were no significant correlations between past (r_τ = .05, *p* = .472,

BF_{01} = 6.46) and current (r_τ = .01, *p* = .911, BF_{01} = 8.56) exposure to violence and *d'*, nor

between past (r_τ = .07, *p* = .308, BF_{01} = 4.14) and current (r_τ = -.03, *p* = .662, BF_{01} = 7.43)

involvement in violence and *d'*, nor between harsh parenting and *d'* ($r_\tau \approx$ .00, *p* = .977, BF_{01} =

8.48) [*Hypothesis 1, individual-level predictions*]. We do not show alpha levels corrected for

multiple comparisons for this set of analyses, because even without corrections, none of the *p*-

values are significant.

Auxiliary analyses: Bias *c*

A Wilcoxon signed rank test showed that the average *c* (*M* = .11, *SE* = .04) also was

significantly higher than zero (*W* = 4044.5, *p* = .018, BF_{10} = 3.34). Separate tests for each group

indicated that *c* did not differ significantly from zero in the community sample (*W* = 871.50, *p* =

.723, BF_{01} = 6.76) [*Hypothesis 2*], but was significantly higher than zero in the student sample

(*W* = 1169.50, *p* = .002, BF_{10} = 44.87) [*Hypothesis 3*]. This indicates that the community sample

was not significantly biased, whereas the student sample did show a no-fight bias.

Welch's *t*-test did not indicate a significant difference in *c* between the community

sample (*M* = .05, *SE* = .06, *Mdn* $\approx$ 0.00) and the student sample (*M* = .17, *SE* = .05, *Mdn* = .14), *t*

= -1.65, *p* = .102, α_{Holm} = .0333, 95% CI [-.26, .02], student BF_{-0} = 1.22 [*Hypothesis 4, group-*

level prediction]. Combining both samples, there were no significant correlations, after correction, between past ($r_\tau = -.13$, $p = .047$, $\alpha_{Holm} = .0250$, $BF_{-0} = 1.94$) and current ($r_\tau = -.15$, $p = .020$, $\alpha_{Holm} = .0167$, $BF_{-0} = 4.12$) exposure to violence and *c*, nor between past ($r_\tau = -.18$, $p = .009$, $\alpha_{Holm} = .0083$, $BF_{-0} = 24.63$) and current ($r_\tau = -.10$, $p = .190$, $\alpha_{Holm} = .0417$, $BF_{0-} = 1.24$) involvement in violence and *c*, nor between harsh parenting and *c* ($r_\tau = -.03$, $p = .604$, $\alpha_{Holm} = .0500$, $BF_{0-} = 5.12$) [*Hypothesis 4, individual-level predictions*].

Exploratory analyses (preregistered)

In the non-preregistered exploratory analyses, we implemented Holm correction (1979) for multiple testing for the two material needs measures (childhood and current). Welch's *t*-tests showed a significant group difference between the two samples in age, as well as childhood and current material needs. The community sample was significantly older ($W = 9.45$, $p < .001$, $BF_{10} > 10000$), and had fewer of their basic material needs met both in their past ($W = -4.68$, $p < .001$, $\alpha_{Holm} = .0500$, $BF_{10} = 2478.14$) and current ($W = -5.83$, $p < .001$, $\alpha_{Holm} = .0250$, $BF_{10} > 10000$) environment than the student sample.

As computed by a Welch's *t*-test for gender and Kendall's tau correlations for our four continuous exploratory predictors (age, life expectancy, childhood and current materials needs), there were no significant relations between the exploratory variables and *d'* (all *p*-values > .05). However, there were significant relations between gender and *c* ($W = 2.03$, $p = .045$, student $BF_{10} = 1.31$), and between age and *c* ($r_\tau = -.14$, $p = .028$, $BF_{10} = 1.61$). Female and younger participants were more likely to have a no-fight bias than male and older participants, respectively.

Exploratory analyses (non-preregistered)

We did not know a priori which of the videos contained enough information to enable

accurate predictions. However, we can re-do our preliminary, core, and auxiliary analyses on

videos for which we do have evidence that they contained enough information: a subset of

videos that people judged better than chance. A non-preregistered exploratory analysis showed

there were statistically significant differences between the videos in the accuracy that people

achieved (see Appendix, Section 3). Separately comparing each video to chance level revealed a

subset of eight videos that people judged better than chance: four fight and four no-fight videos.

The mean proportion of correct responses for this subset was 0.77. Re-doing preliminary

analyses on the subset of eight videos qualitatively changed one result: like the student sample,

the community sample was now more accurate than chance in predicting conflict outcomes. Re-

doing core and auxiliary analyses on this subset produced the same qualitative pattern of non-

significant results as the analyses over all 16 videos. Thus, the community sample was not

significantly biased, whereas the student sample had a no-fight bias.

Discussion

Accuracy

We found no support for our core hypothesis that the community sample would attain

higher *or* lower levels of accuracy in forecasting conflict outcomes than the student sample did.

Overall, participants performed better than chance at forecasting conflict outcomes. However,

this result was driven by the student (low-risk) sample. The community (high-risk) sample

performed at chance. The difference in accuracy between these groups was not significant

[*Hypothesis 1, group-level prediction*].

At the individual level, we also found no support for our core hypothesis that individuals

who have experienced more violence (exposure or active involvement) in their (past or current)

environment attained higher *or* lower accuracy in forecasting conflict outcomes [*Hypothesis 1, individual-level predictions*]. This finding stands out against the broader literature, which tends to report performance deficits on conventional cognitive tests, and mixed evidence on cognitive tasks related to perceiving and memorizing dangerous situations. Null results are rarely reported in this literature. This could be due to publication bias. Another possibility is that, in contrast to conventional cognitive testing (in which people with more experience with violence tend to perform less well on average), our task 'closes the performance gap'. This result would be consistent with the idea that people from harsh environments developmentally adapt to their environments and therefore may show equal rather than impaired performance (Frankenhuis, Young, et al., 2020). A third possibility is that our measurement instruments were not sensitive enough to pick up on such individual differences. Our Bayes Factors indicate the data is more likely under H0 than H1, rather than non-diagnostic. We consider this support for equality of performance.

Bias

We also found no support for our auxiliary hypothesis that the community sample would show a negativity bias by overestimating the number of conflicts that would escalate into fights, relative to the actual number of conflicts that escalated into fights in our task [*Hypothesis 2*]. We did, however, find support for our auxiliary hypothesis that the student sample had a no-fight bias: students underestimated the number of conflicts that escalated into fights [*Hypothesis 3*]. Consistent with some previous research (Laible et al., 2014; Nelson & Crick, 1999), students may have worn 'rose-colored glasses'. The student and community samples may have had different prior expectations that conflicts escalate into fights. Specifically, based on their daily lives, the student sample might have learned that conflicts are less likely to escalate into a fight

than the community sample. In our task, half of the conflicts escalated into a fight. This high rate

may have corresponded more closely to the prior of the community sample than to that of the

student sample. Future research could evaluate this explanation by measuring people's priors

before the task and the structure of the environments that each group is exposed to. For instance,

research could measure the base rates of fights, or the conditional probabilities of conflicts

escalating into fights.

We found no support for our auxiliary hypotheses that the community sample expects

more conflicts to escalate into fights than the student sample [*Hypothesis 4, group-level

prediction*]. At the individual level, we also found no support for our auxiliary hypothesis that

people who have experienced more violence (exposure or active involvement) in their (past or

current) environment had a lower threshold to forecast a conflict would escalate into a fight

[*Hypothesis 4, individual-level predictions*]. This finding is consistent with the results of a recent

study that also found no relation between having more adverse experiences—measured using an

instrument that included items from the same Neighborhood Violence Scale that we used—and

hostile attribution bias in 7- to 10-year-old children (Mesquita & Martins, 2022). Although this

finding is not consistent with the literature on hostile attribution bias (Castro & van Dijk, 2017;

De Castro et al., 2002), the correlation coefficients were in the direction of hostile attribution

bias, with several *p*-values being just above the corrected alpha levels, so it is quite possible that

a higher-powered test of our video paradigm would provide support for hostile attribution bias.

Strengths and limitations

The current study had several strengths. We have conducted a preregistered study of a

novel hypothesis in a hard-to-reach, socioeconomically diverse sample. We used videos of real-

world confrontations, which are ecologically valid (rather than artificial), dynamic (rather than

static), have known outcomes (allowing assessment of accuracy, rather than being ambiguous),

and are likely easier to understand than written or spoken descriptions of (artificial, ambiguous)

situations by an experimenter. However, our study has also important limitations.

First, the test setting differed between our student and community sample. We tested

community participants in a separate room in the community center and students in a cubicle in

the university lab. The main difference between these test settings was that it was noisier in the

room in the community center than in the cubicles in the university lab. In retrospect, it might

have been good to have the experimenters rate the sessions for noisiness to be able to analyze

whether the level of noise was associated with performance. Because it will be difficult for

community participants to visit the lab, future studies may consider testing students in the

community centers (Frankenhuis & Bijlstra, 2018).

Second, the statistical power of our study was likely lower than the 80% we aimed for.

Our preregistered power analysis included an effect size of $d = .5$ for the difference in violence

prediction accuracy across high- and low-risk groups. This effect size was likely too optimistic.

When we conducted our power analysis in 2014, there was less information available than there

is today about how much effect sizes in the literature are inflated (Open Science Collaboration,

2015; Schönbrodt et al., 2013; Szucs et al., 2017). Future research can benefit from adjusting

sample effect sizes for publication bias and uncertainty (Anderson et al., 2017), establishing a

smallest effect size of interest (Lakens, 2022), or using planned sequential sampling designs

(Lakens, 2022; Schott et al., 2019). This work can also draw on recently developed tools that

offer procedures to help determine and justify the sample size (e.g., Kovacs et al., 2022).

Third, whereas our participants were Dutch, the videos were made in countries other than

the Netherlands. One potential explanation of the low rates of accuracy we observed on average,

and of our participants performing at chance on some videos, is that there is a cultural mismatch between the cues indicating whether a conflict in the videos will escalate into a fight, and the cues our participants have learned during their development indicate whether a conflict in the videos will escalate into a fight. Also, the English language in the videos may have been an obstacle for some participants, particularly in the community sample.

Fourth, in the Introduction section, we reasoned that the ability to forecast when conflicts will escalate is more important when violence is relatively common and consequential, in part based on previous studies showing that maltreated children might develop enhanced abilities for detecting angry facial expressions (Masten et al., 2008; Pollak et al., 2009; Pollak & Sinha, 2002). We did not measure maltreatment and do not know how common or rare it was in our sample. Future work may explore whether the ability to forecast conflict outcomes is enhanced in specifically maltreated individuals, and if so, whether this ability is associated with the ability to recognize angry facial expressions among these individuals.

Fifth, we did not know whether participants' priors accurately reflected the properties of their environments. It is likely that people use a combination of prior experience and cues to arrive at their predictions and therefore, their prior experience matters. In our study, we do not know what people's priors are, because we did not measure these priors. However, it is possible that people who experienced more violence have higher priors that conflicts escalate into fights, and people who experienced less violence lower priors. This might explain the findings of rose-colored glasses in our student sample and the lack of hostile attribution bias in our community sample. Future work could examine this explanation by measuring participants' priors–about conflicts escalating into fights–before they watch the study videos. These priors can then be compared with the proportion of 'yes' answers subsequently given in response to the videos.

Finally, it is uncertain how much information about conflict outcomes videos contain. If people judge a conflict outcome better than chance, they may have detected an informative cue, or they may have used a cue that is non-informative (or even a cue associated with the opposite outcome in real life), which lead them to make the right prediction for a particular video. There are examples in other literatures. For instance, a common belief is that people show more signs of nervousness (e.g., gaze aversion, fidgeting) when lying compared with telling the truth, but this is not true; and the opposite might be true when liars actively inhibit signs of nervousness (Vrij et al., 2010). It is similarly challenging to draw inferences when people do 'not judge' a video outcome better than chance, because the absence of evidence is not evidence of absence. For instance, some people may attend to one cue (e.g., intonation), which misled them to make the wrong prediction, and other people to a different cue (e.g., choice of words), which led them to make the right prediction, resulting in chance-level performance across all individuals. Or the majority of people might use both informative and misleading cues. Or people might not detect a cue, even if available. Observational analyses have revealed cues to deception (e.g., involuntary linguistic markers) that untrained people do not detect (ten Brinke & Porter, 2012). Future research may quantify how much information videos contain. Such work can, for instance, use human or computer-automated analyses of real conflicts, some of which escalated and others not, to learn which features predict conflict outcomes. This work may draw on methods already used to analyze surveillance camera recordings of real-life public disputes (Liebst et al., 2021).

Conclusions

We found no support for our core hypothesis that people who have experienced more violence would be better able to forecast whether conflicts will escalate into fights. We also did not find support for our auxiliary hypotheses that the community sample would overestimate the

number of conflicts that would escalate into fights, and that people who have experienced more

violence would be more likely to predict that conflicts will escalate into fights. However, we did

find support for our auxiliary hypothesis that college students would underestimate the number

of conflicts that would escalate into fights. These findings have implications for the literature on

hostile attribution bias, which shows that people with more exposure to violence more likely

interpret the ambiguous actions of others as hostile. Whereas this pattern is typically attributed to

negativity bias (in people with more exposure to violence), it might also reflect rose-colored

glasses (in people with less exposure to violence). Our study could show this because we used

dynamic, real-world stimuli with known outcomes, rather than static, artificial stimuli (e.g.,

vignettes) with unknown outcomes, as previous research has done. Future research on

forecasting conflict outcomes would benefit from measuring people's priors and the statistics of

their actual environments.

References

Albers, C. & Lakens, D. (2018). When power analyses based on pilot data are biased: Inaccurate

effect size estimators and follow-up bias. *Journal of Experimental Social Psychology*, *74*,

187–195. https://doi.org/10.1016/j.jesp.2017.09.004

Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate

statistical power: A method adjusting sample effect sizes for publication bias and

uncertainty. *Psychological Science*, *28*, 1547-1562.

https://doi.org/10.1177/0956797617723724

Aung, T., Goetz, S., Adams, J., McKenna, C., Hess, C., Roytman, S., Cheng, J. T., Zilioli, S., &

Puts, D. (2021). Low fundamental and formant frequencies predict fighting ability among

male mixed martial arts fighters. *Scientific Reports*, *11*, 905.

https://doi.org/10.1038/s41598-020-79408-6

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for

confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*,

255–278. https://doi.org/10.1016/j.jml.2012.11.001

Belsky, J., Spritz, B., & Crnic, K. (1996). Infant attachment security and affective-cognitive

information processing at age 3. *Psychological Science*, *7*, 111–114.

https://doi.org/10.1111/j.1467-9280.1996.tb00339.x

Bjornsdottir, R. T., Alaei, R., & Rule, N. O. (2017). The perceptive proletarian: Subjective social

class predicts interpersonal accuracy. *Journal of Nonverbal Behavior*, *41*, 185–201.

https://doi.org/10.1007/s10919-016-0248-6

Brown, G. S., & White, K. G. (2005). The optimal correction for estimating extreme discriminability. *Behavior Research Methods, 37*, 436–449. https://doi.org/10.3758/BF03192712

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1999). The affect system has parallel and integrative processing components: Form follows function. *Journal of Personality and Social Psychology*, *76*, 839–855. https://doi.org/10.1037/0022-3514.76.5.839

Conger, R. D., Ge, X., Elder, G. H., Lorenz, F. O., & Simons, R. L. (1994). Economic stress, coercive family process, and developmental problems of adolescents. *Child Development*, *65*, 541–561. https://doi.org/10.1111/j.1467-8624.1994.tb00768.x

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information processing mechanisms in children's adjustment. *Psychological Bulletin*, *115*, 74–101. https://doi.org/10.1037/0033-2909.115.1.74

Dang, J., Xiao, S., Zhang, T., Liu, Y., Jiang, B., & Mao, L. (2016). When the poor excel: Poverty facilitates procedural learning. *Scandinavian Journal of Psychology*, *57*, 288–291. https://doi.org/10.1111/sjop.12292

De Castro, B. O., & van Dijk, A. (2017). "It's gonna end up with a fight anyway": Social cognitive processes in children with disruptive behavior disorders. In J. E. Lochman & W. Matthys (Eds.), *Wiley handbook of disruptive and impulse-control disorders* (pp. 237–253). John Wiley & Sons Limited. https://doi.org/10.1002/9781119092254.ch15

De Castro, B. O., Veerman, J. W., Koops, W., Bosch, J. D., Monshouwer, H. J. (2002). Hostile attribution of intent and aggressive behavior: A meta-analysis. *Child Development*, *73*, 916–934. https://doi.org/10.1111/1467-8624.00447

Duncan, G. J., Magnuson, K., & Votruba-Drzal, E. (2017). Moving beyond correlations in assessing the consequences of poverty. *Annual Review of Psychology*, *68*, 413–434. https://doi.org/10.1146/annurev-psych-010416-044224

Eaton, D. K., Kann, L., Kinchen, S., Shanklin, S., Flint, K. H., Hawkins, J., Harris, W. A., Lowry, R., McManus, T., Chyen, D., Whittle, L., Lim, C., & Wechsler, H. (2012). Youth risk behavior surveillance – United States, 2011. *Morbidity and Mortality Weekly Report: Surveillance Summaries, 61,* 1–162. https://www.jstor.org/stable/24806047

Ellis, B. J., Bianchi, J., Griskevicius, V., & Frankenhuis, W. E. (2017). Beyond risk and protective factors: An adaptation-based approach to resilience. *Perspectives on Psychological Science*, *12*, 561–587. https://doi.org/10.1177/1745691617693054

Ellis, B. J., Schlomer, G. L., Tilley, E. H., & Butler, E. A. (2012). Impact of fathers on risky sexual behavior in daughters: A genetically and environmentally controlled sibling study. *Development and Psychopathology*, *24*, 317–332. https://doi.org/10.1017/S095457941100085X

Erdfielder, E., Faul, F., & Buchner, A. (1996). GPower: A general power analysis program. *Behavior Research Methods*, *Instruments, & Computers*, *28*, 1–11. https://doi.org/10.3758/BF03203630

Evans, G. W., Li, D., & Whipple, S. S. (2013). Cumulative risk and child development. *Psychological Bulletin*, *139*, 1342–1396. https://doi.org/10.1037/a0031808

Fields, A., Bloom, P. A., VanTieghem, M., Harmon, C., Choy, T., Camacho, N. L., Gibson, L., Umbach, R., Heleniak, C., & Tottenham, N. (2021). Adaptation in the face of adversity: Decrements and enhancements in children's cognitive control behavior following early

caregiving instability. *Developmental Science*, *24*, e13133.

https://doi.org/10.1111/desc.13133

Finn, E. S., & Bandettini, P. A. (2021). Movie-watching outperforms rest for functional

connectivity-based prediction of behavior. *NeuroImage*, *235*, 117963.

https://doi.org/10.1016/j.neuroimage.2021.117963

Fink, B., Butovskaya, M. L., & Shackelford, T. K. (2019). Assessment of physical strength from

gait: Data from the Maasai of Tanzania. *Biology Letters*, *15*, 20180803.

https://doi.org/10.1098/rsbl.2018.0803

Grandmaster Flash and the Furious Five (1982). The message [Song]. On *The message*. Sugar

Hill Records.

Francis, G. (2014). The frequency of excess success for articles in psychological science.

Psychonomic Bulletin & Review, *21*, 1180–1187. https://doi.org/10.3758/s13423-014-

0601-x

Frankenhuis, W. E., & Bijlstra, G. (2018). Does exposure to hostile environments predict

enhanced emotion detection? *Collabra: Psychology*, *4*, 18.

https://doi.org/10.1525/collabra.127

Frankenhuis, W. E., de Vries, S. A., Bianchi, J., & Ellis, B. J. (2020). Hidden talents in harsh

conditions? A preregistered study of memory and reasoning about social dominance.

Developmental Science, *23*, e12835. https://doi.org/10.1111/desc.12835

Frankenhuis, W. E., & de Weerth, C. (2013). Does early-life exposure to stress shape or impair

cognition? *Current Directions in Psychological Science*, *22*, 407–412.

https://doi.org/10.1177%2F0963721413484324

Frankenhuis, W. E., Roelofs, M., & de Vries, S. A. (2018). Does exposure to social adversity

enhance deception detection ability? *Evolutionary Behavioral Sciences*, *12*, 218–229.

https://doi.org/10.1037/ebs0000103

Frankenhuis, W. E., Young, E. S., & Ellis, B. J. (2020). The hidden talents approach:

Theoretical and methodological challenges. *Trends in Cognitive Sciences*, *24*, 569–581.

https://doi.org/10.1016/j.tics.2020.03.007

Goodman, G. S., Quas, J. A., & Ogle, C. M. (2009). Child maltreatment and memory. *Annual*

Review of Psychology, *61*, 325–351.

https://doi.org/10.1146/annurev.psych.093008.100403

Green, D. M., & Swets, J. A. (1974). *Signal detection theory and psychophysics*. Krieger.

(Original work published 1966)

Hautus, M. J., & Lee, A. (2006). Estimating sensitivity and bias in a yes/no task. *British Journal*

of Mathematical and Statistical Psychology, *59*, 257–273.

https://doi.org/10.1348/000711005X65753

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal*

of Statistics, *6*, 65–70. https://www.jstor.org/stable/4615733

JASP Team (2020). JASP (Version 0.14.0) [Computer software]. https://jasp-stats.org/

Jeffreys, H. (1961). *Theory of probability*. Oxford University Press.

Keiser, R. L. (1979). *The vice lords: Warriors of the streets*. Holt, Rinehart and Winston.

Kovacs, M., van Ravenzwaaij, D., Hoekstra, R., & Aczel, B. (2022). SampleSizePlanner: A tool

to estimate and justify sample size for two-group studies. *Advances in Methods and*

Practices in Psychological Science, *5*, 1-10. https://doi.org/10.1177/25152459211054059

Kraus, M. W., Piff, P. K., Mendoza-Denton, R., Rheinschmidt, M. L., & Keltner, D. (2012).

Social class, solipsism, and contextualism: How the rich are different from the poor.

Psychological Review, *119*, 546–572. https://doi.org/10.1037/a0028756

Kupor, D., Brucks, M. S., & Huang, S.-C. (2019). And the winner is ...? Forecasting the outcome
of others' competitive efforts. *Journal of Personality and Social Psychology, 117*, 500–
521. https://doi.org/10.1037/pspa0000165

Laible, D., McGinley, M., Carlo, G., Augustine, M., & Murphy, T. (2014). Does engaging in
prosocial behavior make children see the world through rose-colored glasses?
Developmental Psychology, 50, 872–880. https://doi.org/10.1037/a0033905

Lakens, D. (2016). Why you don't need to adjust your alpha level for all tests you'll do in your
lifetime. The 20% Statistician. Retrieved from:
http://daniellakens.blogspot.ch/2016/02/why-you-dont-need-toadjust-you-alpha.html

Lakens, D. (2022). Sample size justification. *Collabra: Psychology*, *8*, 33267.
https://doi.org/10.1525/collabra.33267

Lakens, D., & Etz, A. J. (2017). Too true to be bad: When sets of studies with significant and
non-significant findings are probably true. *Social Psychological and Personality Science*,
8, 875–881. https://doi.org/10.1177/1948550617693058

Lane, S. M., & Briffa, M. (2020). Perceived and actual fighting ability: Determinants of success
by decision, knockout or submission in human combat sports. *Biology Letters*, *16*,
20200443. https://doi.org/10.1098/rsbl.2020.0443

Leve, L. D., Harold, G. T., Chamberlain, P., Landsverk, J. A., Fisher, P. A., & Vostanis, P.
(2012). Practitioner review: Children in foster care–vulnerabilities and evidence-based
interventions that promote resilience processes. *Journal of Child Psychology and
Psychiatry*, *53*, 1197–1211. https://doi.org/10.1111/j.1469-7610.2012.02594.x

Liebst, L. S., Philpot, R., Levine, M., & Lindegaard, M. R. (2021). Cross-national CCTV footage
shows low victimization risk for bystander interveners in public conflicts. *Psychology of*

Violence, *11*, 11-18. https://doi.org/10.1037/vio0000299

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.).
    Lawrence Erlbaum.

Masten, C. L., Guyer, A. E., Hodgdon, H. B., McClure, E. B., Charney, D. S., Ernst, M.,
    Kaufman, J., Pine, D. S., & Monk, C. S. (2008). Recognition of facial emotions among
    maltreated children with high rates of post-traumatic stress disorder. *Child Abuse &
    Neglect*, *32*, 139–153. https://doi.org/10.1016/j.chiabu.2007.09.006

Mesquita, S., & Martins, P. C. (2022). Social information-processing in children with adverse
    experiences. *Trauma Care*, *2*, 162-173. https://doi.org/10.3390/traumacare2020013

Mittal, C., Griskevicius, V., Simpson, J. A., Sung, S., & Young, E. S. (2015). Cognitive
    adaptations to stressful environments: When childhood adversity enhances adult
    executive function. *Journal of Personality and Social Psychology*, *109*, 604–621.
    https://doi.org/10.1037/pspi0000028

Nelson, D. A., & Crick, N. R. (1999). Rose-colored glasses: Examining the social information-
    processing of prosocial young adolescents. *Journal of Early Adolescence*, *19*, 17–38.
    https://doi.org/10.1177/0272431699019001002

Nweze, T., Nwoke, M. B., Nwufo, J. I., Aniekwu, R. I., & Lange, F. (2020). Working for the
    future: Parentally deprived Nigerian children have enhanced working memory
    ability. *Journal of Child Psychology and Psychiatry*, *62*, 280–288.
    https://doi.org/10.1111/jcpp.13241

Open Science Collaboration (2015). Estimating the reproducibility of psychological science.
    Science, *349*, aac4716. https://doi.org/10.1126/science.aac4716

Pollak, S. D. (2008). Mechanisms linking early experience and the emergence of emotions:

Illustrations from the study of maltreated children. *Current Directions in Psychological Science*, *17*, 370–375. https://doi.org/10.1111/j.1467-8721.2008.00608.x

Pollak, S. D., Messner, M., Kistler, D. J., & Cohn, J. F. (2009). Development of perceptual expertise in emotion recognition. *Cognition*, *110*, 242–247. https://doi.org/10.1016/j.cognition.2008.10.010

Pollak, S. D., & Sinha, P. (2002). Effects of early experience on children's recognition of facial displays of emotion. *Developmental Psychology*, *38*, 784–791. https://doi.org/10.1037/0012-1649.38.5.784

R Core Team. (2020). *R: A language and environment for statistical computing* (Version 4.0.3) [Computer software]. R Foundation for Statistical Computing. https://www.Rproject.org/

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*, 225–237. https://doi.org/10.3758/PBR.16.2.225

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*, 296–320. https://doi.org/10.1207/S15327957PSPR0504_2

Schimmack, U. (2012). The ironic effect of significant results on the credibility of multiple-study articles. *Psychological Methods*, *17*, 551–566. https://doi.org/10.1037/a0029487

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize?. *Journal of Research in Personality*, *47*, 609-612. https://doi.org/10.1016/j.jrp.2013.05.009

Schott, E., Rhemtulla, M., & Byers-Heinlein, K. (2019). Should I test more babies? Solutions for

    transparent data peeking. *Infant Behavior and Development*, *54*, 166–176.

    https://doi.org/10.1016/j.infbeh.2018.09.010

Shakur, S. (2007). *Monster: The autobiography of an LA gang member*. Grove Press.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory:

    Applications to dementia and amnesia. *Journal of Experimental Psychology: General*,

    117, 34–50. https://doi.org/10.1037/0096-3445.117.1.34

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection measures. *Behavior*

    Research Methods, Instruments, & Computers, 31, 137–149.

    https://doi.org/10.3758/BF03207704

Szucs, D., & Ioannidis, J. P. (2017). Empirical assessment of published effect sizes and power in

    the recent cognitive neuroscience and psychology literature. *PLoS biology*, *15*, e2000797.

    https://doi.org/10.1371/journal.pbio.3001151

Ten Brinke, L., & Porter, S. (2012). Cry me a river: Identifying the behavioral consequences of

    extremely high-stakes interpersonal deception. *Law and Human Behavior*, *36*, 469.

    https://doi.org/10.1037/h0093929

Vandenbroucke, L., Verschueren, K., Ceulemans, E., De Smedt, B., De Roover, K., & Baeyens,

    D. (2016). Family demographic profiles and their relationship with the quality of

    executive functioning subcomponents in kindergarten. *British Journal of Developmental*

    Psychology, *34*, 226–244. https://doi.org/10.1111/bjdp.12127

Venkatesh, S. (2008). *Gang leader for a day: A rogue sociologist takes to the streets*. Penguin

    Press.

Vrij, A., Granhag, P. A., & Porter, S. (2010). Pitfalls and opportunities in nonverbal and verbal

lie detection. *Psychological Science in the Public Interest*, *11*, 89-121.

https://doi.org/10.1177/1529100610390861

Watkins, I. J., & Martire, K. A. (2015). Generalized linear mixed models for deception research:

Avoiding problematic data aggregation. *Psychology, Crime & Law*, *21*, 821–835.

https://doi.org/10.1080/1068316X.2015.1054384

Young, E. S., Griskevicius, V., Simpson, J. A., Waters, T. E., & Mittal, C. (2018). Can an

unpredictable childhood environment enhance working memory? Testing the sensitized-

specialization hypothesis. *Journal of Personality and Social Psychology*, *114*, 891–908.

https://doi.org/10.1037/pspi0000124

Author contributions

Contributed to conception and design: WEF, JB

Contributed to acquisition of data: WEF, JB

Contributed to analysis and interpretation of data: WEF, ELW, SAV, MZ, JB

Drafted and/or revised the article: WEF, ELW, SAV

Approved the submitted version for publication: WEF, ELW, SAV, MZ, JB

Competing interests

On behalf of all authors, I, WEF, declare to not have any competing interests that might (be perceived to) influence the interpretation of this article.

Supplemental material

This article is accompanied by an appendix that provides box plots for our dependent variables (accuracy and bias), non-preregistered Generalized Linear Mixed Models of core and auxiliary

analyses, and non-preregistered core and auxiliary analyses of the subset of eight videos for which there is evidence that these contained information.

Data accessibility statement

Our materials, data, and analysis script are available at:

https://osf.io/2rt95/?view_only=006fc220639e4e16bbb4a2e93f824c87. We do not have copyright permission to upload the videos we used under a permalink. Therefore, we can only share these videos upon request.

Figure title and legends

Our appendix includes Figure S1 captioned "Boxplots showing the distribution of values of d' in both risk groups".

Our appendix includes Figure S2 captioned "Figure S2. Boxplots showing the distribution of values of c in both risk groups".

Our appendix includes Figure S3 captioned "The mean proportion of correct answers per video. The dotted horizontal lines represent 1 and 2 standard deviations as expected based on a binomial distribution with a probability of 0.5, with our final sample size of 127 (i.e., 127 trials per video). The proportions of correct answers that significantly differed from chance (as based on exact binomial tests) were those of video 1, 2, 4, 6, 7, 8, 10, and 11, so these were the videos we included in our subset. The mean proportion of correct responses for this subset was 0.77".