

Research Report

Fear Extinction to an Out-Group Face

The Role of Target Gender

Carlos David Navarrete,¹ Andreas Olsson,² Arnold K. Ho,³ Wendy Berry Mendes,³ Lotte Thomsen,³ and James Sidanius³

¹Michigan State University, ²Columbia University, and ³Harvard University

ABSTRACT—Conditioning studies on humans and other primates show that fear responses acquired toward danger-relevant stimuli, such as snakes, resist extinction, whereas responses toward danger-irrelevant stimuli, such as birds, are more readily extinguished. Similar evolved biases may extend to human groups, as recent research demonstrates that a conditioned fear response to faces of persons of a social out-group resists extinction, whereas fear toward a social in-group is more readily extinguished. Here, we provide an important extension to previous work by demonstrating that this fear-extinction bias occurs solely when the exemplars are male. These results underscore the importance of considering how gender of the target stimulus affects psychological and physiological responses to out-group threat.

Research in prepared learning demonstrates that fear responses conditioned to danger-relevant stimuli, such as spiders and snakes, resist extinction, whereas responses toward danger-irrelevant stimuli, such as birds or butterflies, are more readily extinguished (Öhman & Mineka, 2001). Such biases in fear conditioning are said to be “prepared” in domains toward which a species has had sufficient exposure over time for natural selection to affect the neural circuitry underlying associative learning mechanisms. These mechanisms can then give rise to functional behavioral changes that emerge over the lifetime of the individual, such as maintaining fear toward dangerous stimuli to which one has had a negative experience, thereby avoiding future harm (Seligman, 1971).

Though the fear system underlying prepared learning may be useful under some circumstances, it may be at the root of some

persistent social problems affecting modern societies—including xenophobia. Previous studies have observed that race bias and fear learning rely on overlapping neural systems (e.g., Phelps et al., 2000), suggesting a shared mechanistic link between the two, and thus the potential to use a fear conditioning paradigm to investigate prepared fear learning in an intergroup context. Recently, Olsson, Ebert, Banaji, and Phelps (2005) reported that conditioned fear toward facial displays of male individuals belonging to a racial group other than one’s own resists extinction, whereas fear toward faces of one’s own racial group does not. Their results held for both White and Black American research participants toward White and Black out-group targets and were unrelated to participants’ measured level of negative attitudes against the racial out-group. The sole behavioral variable found to be associated with a reduction in conditioned fear was participants’ history of intimate interracial contact. Such findings suggest that, although the mechanisms underlying prepared learning of out-group fear may be due both to evolved social categorization mechanisms and life-history experiences, the psychological system dedicated to fear learning of out-groups may operate orthogonally to those processing areas that manage socially transmitted stereotypes and attitudes.

THE PRESENT STUDY

Given the knowledge that males have historically been the primary agents of intergroup aggression in humans (Daly & Wilson, 1988; Keeley, 1996; Kelly, 2005; Wrangham & Peterson, 1996) and that the potential for harm present in the stimulus prepares the fear system for functionally specialized behavioral outcomes such as fight, freeze, or flight (Öhman & Mineka, 2001; Seligman, 1971), we predicted that the fear-extinction bias between in-group and out-group faces found by Olsson et al. (2005) would occur solely when the exemplars are male. That is, in a fear-conditioning experiment, we tested the prediction that conditioned fear to the

Address correspondence to Carlos David Navarrete, Department of Psychology, Michigan State University, East Lansing, MI 48824, e-mail: cdn@msu.edu.

face of a male out-group target would resist extinction, but that conditioned fear toward the face of a female out-group target, or toward in-group males and females, would be readily extinguished. We also report an analysis that explored the extent to which the persistence of conditioned fear to the face of an out-group male is tied to socially transmitted stereotypes or to a history of close, intergroup contact.

METHOD

Participants

Study participants were 165 White and 35 Black United States citizens from the psychology study pools at Harvard University ($n = 85$) and Michigan State University ($n = 115$). Volunteers were composed of students, university staff, and community members who were paid \$20 or given course credit to participate in “a study that explores the mind-body connection in response to social groups.” We used the widely accepted exclusionary criteria adopted by Olsson et al. (2005): Data from 33 participants were excluded from the analysis because of technical problems ($n = 5$), lack of a skin conductance response ($n = 13$), or failure to acquire a conditioned response to at least one of the two reinforced conditioned stimuli during acquisition ($n = 11$). After beginning the procedure, 4 participants elected to discontinue participation, leaving an analyzable sample consisting of 139 White and 28 Black American participants (98 females, 69 males; age range = 18–61 years, mean age = 21.8 years, $SD = 7.1$ years). Stimuli and experimental protocol were identical across both samples.

Pretest Procedure

Upon arrival, participants completed pretest measures that included the measures listed below. Descriptive statistics for each pretest measure are reported in Table 1.

Explicit Race Bias

Explicit race bias was measured using Attitudes Toward Blacks scale (Brigham, 1993). The scale included items like “Generally, Blacks are not as smart as Whites” and “It is likely that Blacks will bring violence to neighborhoods when they move in.” Black American participants completed the measure with the word “Whites” substituted for “Blacks.” Items were assessed on a 7-point response scale anchored at both poles (1 = *strongly agree*, 7 = *strongly disagree*). Cronbach’s test for scale reliability was satisfactory ($\alpha = .88$).

Implicit Race Bias

Implicit bias was measured via the Implicit Association Test (IAT), a method that uses reaction times in pairing certain concepts into related categories (Greenwald, McGhee, & Schwartz, 1998). The method is said to reveal bias in the strength of the automatic associations between pairs of those concepts in memory, such as “good/White” versus “bad/Black.” Two IATs were used: stereotypic and evaluative race bias. One measured the

TABLE 1

Pretest Measures and Their Correlations With Extinction Bias in the Male-Target Condition

Bias	<i>n</i>	<i>M</i>	<i>SD</i>	Correlation (<i>r</i>) with extinction bias
Extinction bias	84	0.08	0.25	—
Explicit race bias	76	2.45	−0.90	−.02
Implicit race bias, evaluative	82	0.36	0.35	−.19
Explicit race bias, stereotype	84	0.26	0.38	.11
Contact	84	1.62	0.96	−.24*

* $p < .05$.

biased associations of physical and mental concepts stereotypically applied to White and Black Americans (e.g., *math, brainy, athletic, strong*). The other measured biased associations of evaluative concepts not typically associated with racial stereotypes (e.g., *joy, love, agony, horrible*) but with high affective valence (Amodio & Devine, 2006).

Out-Group Contact

Contact items measured the number of White and Black friends, acquaintances, and romantic partners had by participants. Past interracial contact was coded such that greater values indicated more out-group relative to in-group contact. The relative-contact measure was created by subtracting the number of in-group contacts from the number of out-group contacts (Olsson et al., 2005).

Conditioning Procedure

After the pretest, participants underwent a delayed fear-conditioning protocol, where a conditioned response was engendered to four categories of conditioned stimuli via electric shock and noise. Conditioned stimuli were composed of images of White and Black American male and female faces that appeared on a computer screen.

Before the procedure, skin conductance electrodes were attached to the second and fourth distal phalanges of the left hand, and shock electrodes were attached to the right wrist. Shock amplitude was then assessed by the participant as “uncomfortable, but not painful” by a work-up procedure.

During fear conditioning, participants were presented with Black and White facial images such that the target’s racial group was experimentally manipulated within subjects. Half the participants were exposed to male faces only, and the other half were exposed solely to female faces, such that the gender of the target exemplar was manipulated between subjects.¹ Each stimulus was presented once per trial across three learning

¹Male faces were identical to those used by Olsson et al. (2005). Female faces were from Models 7F, 9F, 11F, and 13F from the MacBrain Face Stimulus set, overseen by Nim Tottenham and supported by the John D. and Catherine T. MacArthur Foundation Research Network on Early Experience and Brain Development. All faces had neutral expressions.

phases: habituation (three trials), acquisition (five trials), and extinction (six trials). During acquisition, one image from each stimulus category (the reinforced conditioned stimulus, CS+) was paired with an aversive outcome (the unconditioned stimulus, UCS), whereas the other image (the unreinforced conditioned stimulus, CS-) was presented without the UCS. An electrical shock simultaneously paired with a short burst of white noise (90 dB) together constituted the UCS. Each conditioned stimulus was presented for 6 s and coterminated with the UCS or not (CS+/-); this was followed by an interstimulus interval of 12 to 14 s. During the extinction phase that followed, stimuli were presented without the UCS.

RESULTS

Conditioned-Response Scoring

Skin conductance responses (SCRs) were recorded during the presentation of each stimulus (0.5–4.5 s after onset). Only the largest SCRs (minimal response criterion of 0.02 μ S) were used, and raw values were root-squared to normalize the data. The conditioned response (CR) was operationalized as the mean differential SCR between the CS+ and the CS- from the same category, so as to minimize preexisting differences as a potential confound. Habituation means comprised CRs to the first four presentations, acquisition means comprised CRs to the five presentations after the first trial of the CS+ paired with the UCS (i.e., presentations 5–9 of each CS), and extinction means were based on the CRs to the last five presentations of each CS (i.e., presentations 10–14). Data were analyzed for participants whose acquisition mean for at least one of the two CS+ was greater than zero.

Data Analyses

Descriptive statistics for all conditioning phases appear in Table 2. Consistent with our expectations, the mean CRs did not differ significantly from zero during extinction for exemplars representing in-group females, $t(82) = 1.27, p = .21$; out-group females, $t(82) < 1, p = .55$; or in-group males, $t(83) < 1, p = .39$, whereas the CR to an out-group male exemplar was resistant to extinction, $t(83) = 3.48, p = .0008$. A 2×2 mixed effects analysis of variance, in which target gender was the between-subjects effect and target race was the within-subjects effect, revealed a significant two-way interaction, $F(1, 166) = 4.74, \eta^2 = .47, p = .03$, demonstrating the extinction bias between in-group and out-group targets was greater when the exemplars were male (Fig. 1).

When entered as covariates in this model, dummy variables for subject race, subject gender, and sample location yielded a significant main effect in fear extinction for race ($p < .01$), such that Black subjects showed greater resistance to extinction than White subjects. However, the Target Race \times Target Gender interaction term was unaffected by this main effect, and these variables did not significantly interact with target gender or target group.

TABLE 2

Descriptive Statistics for Conditioned Response to Male and Female Exemplars, by Group and Trial Type

Group and trial type	Skin conductance (μ S)		
	<i>M</i>	<i>SD</i>	95% confidence interval
Female exemplars ($n = 83$)			
In-group			
Habituation	0.02	0.24	-0.04–0.07
Acquisition	0.22	0.24	0.16–0.27
Extinction	0.03	0.20	-0.02–0.07
Out-group			
Habituation	0.01	0.20	-0.03–0.06
Acquisition	0.23	0.24	0.18–0.28
Extinction	0.02	0.20	-0.02–0.06
Male exemplars ($n = 84$)			
In-group			
Habituation	0.00	0.22	-0.05–0.04
Acquisition	0.30	0.36	0.22–0.38
Extinction	0.01	0.20	-0.03–0.06
Out-group			
Habituation	-0.03	0.26	-0.08–0.03
Acquisition	0.28	0.29	0.22–0.35
Extinction	0.09	0.24	0.04–0.14

We then explored the extent to which individual differences in interracial contact or racial attitudes assessed in a pretest might be related to out-group bias in fear extinction between in-group and out-group male targets. Because CRs for both in-group and out-group were correlated, we computed a measure of extinction bias by subtracting in-group face CRs from out-group face CRs, with higher values indicative of greater out-group bias in resistance to extinction of conditioned fear. The analysis was

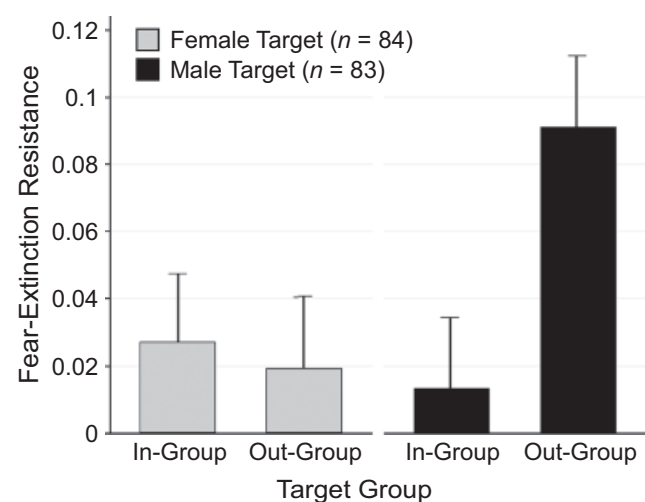


Fig. 1. Fear-extinction resistance by target gender and target group. Higher values denote greater resistance to extinction of a conditioned response, as measured by skin conductance (in microsiemens). Zero values denote complete extinction, and error bars indicate standard errors.

restricted to observations in the male target condition across the extinction phase. Bivariate correlations among extinction bias and explicit race bias, implicit race bias, and interracial contact, revealed that intergroup contact was the sole significant correlate (Table 1). That is, the CR to out-group males was extinguished more readily in individuals with greater intergroup contact.

DISCUSSION

In a fear-conditioning paradigm in which a fear response was engendered toward male and female faces of White and Black Americans, we found that social out-group targets served as prepared stimuli, but only when the exemplar was male. This finding suggests that the resistance to extinction of conditioned fear toward an out-group target, demonstrated by Olsson et al. (2005), is not likely to be caused by a psychological system that operates merely on in-group and out-group categorical distinctions, but one that uses gender categories as well. Further research is needed to explore whether this distinction occurs because human male facial physiognomy might naturally connote anger (Becker, Kenrick, Neuberg, Blackwell, & Smith, 2007; Dimberg & Öhman, 1996), or because mere gender categorization can act as a heuristic cue for the potential for danger the stimulus has posed over evolutionary time. In either case, these results are consistent with the notion that intergroup conflict perpetrated by male aggressors has been common enough over our evolutionary history to have allowed natural selection sufficient time to have shaped the neural circuitry underlying contemporary expressions of intergroup bias.

Consistent with Olsson et al. (2005), we found no evidence that the explicit endorsement or implicit internalization of racial stereotypes is related to male-targeted race bias in fear extinction. Taken together with previous findings demonstrating that such prepared learning does not extend to other culturally defined fear-relevant stimuli, such as visual images of broken electrical outlets and firearms (e.g., Cook, Hodes, & Lang, 1986), this suggests that our results are not likely to be due solely to the salience of socially transmitted, semantic information reinforcing negative associations with the prepared stimuli. However, our replication of the finding that a reduction in the prepared effect is penetrable via close, intimate contact (Olsson et al., 2005) suggests that developmental triggers or dispositional factors that emerge early in life may facilitate the reduction of out-group fear, thereby increasing out-group contact.

Evidence that infants evince gender bias in stranger anxiety (Freedman, 1961) suggests that the psychological system for prepared fear underlying our results is sensitive to sex-differentiated patterns of aggression even in the early months of human development. We consider our results an early step in exploring why male-directed out-group fear emerges so early and persists so reliably. We are hopeful that investigations using more detailed individual-difference measures known to be associated with neurophysiological markers of anxiety and race bias may provide

further insight into the evolved psychological architecture underlying the human predisposition for xenophobia.

Acknowledgments—This material is based on work supported by National Science Foundation Grant SES-0409798. We thank Marilyn Brewer, Tony Greenwald, Dan Fessler, Diana Fleischman, Norb Kerr, and Andreas Wilke for their helpful comments on an earlier version of the manuscript, and Alissa Cardone, Katie LaRoche, Katie Heikkinen, Ben Asher, Kierstin Lorence, Joe Sbar, Jnanna David, and Nicole Boucher for lab assistance.

REFERENCES

- Amodio, D.M., & Devine, P.G. (2006). Stereotyping and evaluation in implicit race bias: Evidence for independent constructs and unique effects on behavior. *Journal of Personality and Social Psychology, 91*, 652–661.
- Becker, D.V., Kenrick, D.T., Neuberg, S.L., Blackwell, K.C., & Smith, D.M. (2007). The confounded nature of angry men and happy women. *Journal of Personality and Social Psychology, 92*, 179–190.
- Brigham, J.C. (1993). College students' racial attitudes. *Journal of Applied Social Psychology, 23*, 1933–1967.
- Cook, E.W., III, Hodes, R.L., & Lang, P.J. (1986). Preparedness and phobias: Effects of stimulus content on human visceral conditioning. *Journal of Abnormal Psychology, 95*, 195–207.
- Daly, M., & Wilson, M. (1988). *Homicide*. Hawthorne, NY: Aldine.
- Dimberg, U., & Öhman, A. (1996). Behold the wrath: Psychophysiological responses to facial stimuli. *Motivation & Emotion, 20*, 149–182.
- Freedman, D.G. (1961). The infant's fear of strangers and the flight response. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 2*, 242–248.
- Greenwald, A.G., McGhee, D.E., & Schwartz, J.L.K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*, 1464–1480.
- Keeley, L.H. (1996). *War before civilization: The myth of the peaceful savage*. New York: Oxford University Press.
- Kelly, R.C. (2005). The evolution of lethal intergroup violence. *Proceedings of the National Academy of Sciences, USA, 102*, 15294–15298.
- Öhman, A., & Mineka, S. (2001). Fear, phobias and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review, 108*, 483–522.
- Olsson, A., Ebert, J.P., Banaji, M.R., & Phelps, E.A. (2005). The role of social groups in the persistence of learned fear. *Science, 309*, 785–787.
- Phelps, E.A., O'Connor, K.J., Cunningham, W.A., Funayama, E.S., Gatenby, J.C., Gore, J.C., & Banaji, M.R. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience, 12*, 729–738.
- Seligman, M. (1971). Phobias and preparedness. *Behavior Therapy, 2*, 307–321.
- Wrangham, R.W., & Peterson, D. (1996). *Demonic males: Apes and the origins of human violence*. Boston: Houghton Mifflin.

(RECEIVED 5/23/08; REVISION ACCEPTED 7/31/08)