User Name: llkong@AG.GOV.TT
Date and Time: Wednesday 4 June 2025 16:32 EDT
Job Number: 254538267

Document (1)

1. Estimating the perceived threat threshold for acting in self-defense

Search Terms: stand-your-ground

Search Type: Natural Language

Narrowed by:

Content Type	Narrowed by
Books & Journals	-None-

Estimating the perceived threat threshold for acting in self-defense

Law Probability and Risk (2024) 23 (1): 1

1 March 2024

Law, Probability and Risk > 2024 > Issue 1, March > Articles

Law Probability and Risk

Aili Qiao

Department of Psychology, University of Southern California, Los Angeles, CA 90089, United States

Department of Psychology, University of Southern California, 3620 S. McClintock Ave., Seeley G. Mudd Building, Room 501, Los Angeles, CA 90089, United States. E-mail: aqiao@usc.edu

Richard S. John

Department of Psychology, University of Southern California, Los Angeles, CA 90089, United States

Abstract: We formulate the decision to engage in self-defense within a Bayesian framework and report an experiment assessing the influence of situational cues and individual difference variables on perceptions of threat and intentions to act in self-defense. US adult respondents (N = 235) read four ambiguous threat vignettes involving situations in which the respondent is alone in a high-or low-crime area and is approached by a male stranger exhibiting either high or low immediate threat cues. For each vignette, respondents indicated a probability that the stranger intends to harm and a binary intention to either engage in proactive self-defense or not. The extent to which manipulated proximal and distal cues influenced threat perception probabilities was moderated by both respondents' gender and political orientation. Threat perception probabilities mediate the influence of both proximal and distal cues on the intention to engage in proactive self-defense. Even odds thresholds of perceived threat probability to engage in self-defense ranged from 0.62 to 0.65 across the four scenarios, suggesting that failing to react to an actual threat is perceived as two to three times more costly than that of mistaken self-defense.

Keywords: proactive self-defense; Bayes theorem; Protection Motivation Theory; Protective Action Decision Model; political orientation.

1. Introduction

It is estimated that self-defense has the highest rate of usage among all criminal defenses, and this estimate is not surprising given that self-defense has the highest success rate, as well as impact on attenuating sentences, compared to all other criminal defenses (Cohen, Johnson, and Henley 1993; Valdes 2005). At its core, a successful self-defense defense has three elements. First, the defender must honestly believe that the perceived threat from the possible attacker is imminent (Nourse 2001). Second, the defender must honestly believe that the self-defense act chosen is necessary in the context of the perceived threat (Oleson and Darley 1999). Third, the defender must have reasonable grounds for their beliefs (Kim 2008). Although some have argued for an objective standard for

necessity (McMahan 1994; Fletcher 1996; Fontaine 2010), the defender's beliefs need not be objectively true, for example, it could very well be that the attacker's gun was unloaded or did not intend to harm the defender.

If the defender's beliefs about the perceived threat and necessity to act are deemed reasonable in a court of law, the defender can successfully claim self-defense (Heller 1998; Ferzan 2005; Wallerstein 2005). These three fundamental elements compose the essence of the Model Penal Code's general guideline for ruling on self-defense cases: “the use of force upon or toward another person is justifiable when the actor believes that such force is immediately necessary for the purpose of protecting himself against the use of unlawful force by such other person on the present occasion (MPC, §3.04; American Law Institute, 1962).”

However, states across the USA have varying interpretations of what “justifiable force under self-defense” exactly entails. Laws in twenty-eight states, including Florida, Michigan, and North Carolina, have stand-your-ground by statute, allowing defenders to protect themselves in

any lawfully present place. Eight states, including California, Colorado, and Vermont, do not have stand-your-ground by statute; instead, they either implement stand-your-ground via jury instructions or allow stand-your-ground case law/precedent in court. Finally, laws in eleven states, including Delaware, Maine, and Minnesota impose duty-to-retreat in public spaces, where defenders cannot employ force unless they have exhausted all options to retreat. Overall, there is no consensus across the USA as to the circumstances under which self-defense is “justifiable,” even though the national crime data reported by the Federal Bureau of Investigation (FBI) suggest that there are a total of 300–400 cases of justifiable homicide every year committed by private citizens (FBI 2020).

Figure 1 provides a schematic overview of the self-defense defense, demonstrating the essential role of the defendant's perception of an imminent attack. The jury must consider the defendant's perception of the situation and compare it to other situational evidence to assess the veracity of the defendant's claimed perceptions. In addition, the jury must consider the defendant's belief that self-defense is necessary to counter the perceived threat. Under some state laws, the defendant's alternatives to retreat must be considered in judging whether an act of self-defense is warranted. Finally, the jury must decide whether the defendant's actions were justifiable by combining their appraisal of the defendant's perceived threat with their evaluation of the necessity for action on the defendant's part to neutralize the perceived threat.

Therefore, as a first step to addressing the controversy around state laws and self-defense success rate, it is crucial to understand how individuals perceive threats leading to self-defense against a possible attacker. In our research, we introduce a Bayesian framework for modeling threat perception in self-defense acts. Using this framework, we seek to examine how individuals' perceptions of threat and its associated prior/posterior probabilities inform their decision to engage in self-defense acts.

1.1 Theories of self-protective behavior

Self-protective behavior involves engaging in protective measures in anticipation of a future threat. Self-protective behaviors are extensively studied in fields such as health (e.g. exercising, going to the doctors, purchasing health insurance, etc.) and natural disasters (e.g. disaster preparation, early evacuation, buying home insurance, etc.). In the context of violent crimes, self-protective behavior—or self-defense—is typically defined as defending oneself against a possible attacker using firearms, knives, or physical weapons. However, the National Crime Victimization Survey data released in 2020 reported that only 8.6% of violent crime victims attempted physical self-defense.

Figure 1. Schematic overview of the self-defense defense.

A few theories address threat perception in self-protective behavior, notably Ronald Rogers' Protection Motivation Theory (PMT Rogers, 1975; Maddux & Rogers, 1983). Figure 2 presents a conceptual framework for PMT, which posits that threat and coping appraisal affect one's attitude and prediction of a potential threat. Rogers suggests that whether one engages in self-protective behavior relies on appraisals of the presented threat cues and the efficacy and potential costs of available coping methods. In a later investigation, Rogers (1985) concludes that threat appraisal is the sum of (1) one's perception of threat noxiousness and (2) the probability of threat occurrence. Rogers suggests that individuals are more motivated to engage in protective actions when an event is perceived as harmful and likely to occur. PMT has been applied successfully in a variety of contexts to account for self-protective behavior (Floyd et al. 2000; Milne et al. 2000).

Another relevant theory that addresses threat perception is Lindell and Perry's Protective Action Decision Model (PADM; 2004, 2012). As indicated in Fig. 3, PADM suggests a sequential process of PADM, which begins with the reception of threat cues that elicit assessments of (1) external threat, (2) protective actions, and (3) relevant stakeholders, resulting in the final self-protective behavior. The PADM framework has been applied extensively in contexts involving decisions related to natural disasters (Heath et al. 2018) as well as in contexts involving responses to aggressive actions (Wang et al. 2020). The authors suggest that some critical threat assessment factors include perceived probability, potential consequences, feelings of dread, and uncertainty.

Though both theories address the cognitive process of threat assessment that leads to protective behavior, PMT describes a person's *prior* beliefs about the noxiousness/severity and the probability of a threat occurrence—for example, how one generally estimates an adverse event's chances and outcomes. On the other hand, PADM focuses on assessing the situation as the agent receives *immediate* threat cues—for example, how one navigates possible responses given the information provided at a given moment. Humans likely assess threats in a more holistic view: that is incorporating both their prior beliefs about the threat context and the perception of immediate threat cues. Therefore, we conceptualize a threat perception model to include both.

1.2 A Bayesian framework for preemptive self-defense

A schematic overview of a Bayesian framework for preemptive self-defense action is presented in Fig. 4. Normatively, an individual's decision to engage in preemptive self-defense should depend on a comparison of the proximal threat cues with the threshold for taking action. The

Figure 2. Conceptual framework of PMT.

Figure 3. Conceptual framework of the PADM.

Figure 4. Schematic of Bayesian framework for preemptive self-defense action.

action threshold normatively depends on prior (distal) knowledge of the context and the relative cost of errors, that is engaging in self-defensive action when there is no actual threat versus not engaging in self-defense when there is an actual threat. Distal cues are distinct from the immediate threat and relate to the context of time and place. Distal cues represent the defendant's prior beliefs about the likelihood of a threat in that particular place and time. Proximal cues relate directly to perceived immediate threat signals, such as the potential attacker's appearance and behavior, both verbal and physical. Possession and brandishing of a weapon would also be considered proximal cues. Proximal threat signals and distal prior knowledge may also be combined normatively into an assessment of the probability of an imminent attack.

Previous research has demonstrated that distal cues are often ignored in favor of more salient proximal cues

(Tversky and Kahneman 1974; Lyon and Slovic 1976; Bar-Hillel 1980). This early research suggested that uncertainty appraisal may focus primarily on the immediate proximal cues and neglect to consider distal base-rate cues. Christensen-Szalanskiand Beach (1982) reported that experiential base rates are more likely to be perceived as relevant and utilized in appraisals of uncertainty. This finding has been replicated and extended over the last 40 years

(e.g. Betsch et al. 1998; Goodie and Fantino 1999; Welsh and Navarro 2012; Newman et al. 2017). Koehler (1996) suggested that the use or neglect of base rates is contingent on several factors, and his “it depends” conclusions have been reported in several more recent empirical studies (Lovett and Schunn 1999; Turpin et al. 2020; Pilditch et al. 2021).

One important general finding is that base-rate neglect is moderated by the degree to which the information is perceived to have causal relevance to the uncertainty appraisal (Sloman, 2003; Krynski and Tenenbaum 2007). Previous research suggests that distal cues in the form of base rates will be utilized if they are perceived as causally relevant to the appraisal of an imminent threat of attack. Neighborhood crime rates are likely to be perceived as causally relevant for the appraisal of the threat of an imminent attack. Normatively, it is desirable to utilize accurate base rates. Ignoring base rates is a fallacy that can result in suboptimal responses to observed proximal threat cues. Base rates, or prior odds, are subjective in the sense that they are based on each individual's previous experience and beliefs. Decisions to act in self-defense should include consideration of these beliefs in the form of the individual's prior odds of an imminent attack, which may or may not be an accurate reflection of reality.

The defendant's posterior odds of an imminent attack (Ha), following observation of proximal threat cues (T), should be the product of the prior odds of an attack, based on prior contextual knowledge, and the likelihood ratio (Bayes Factor) representing the ratio of the probability of the observed proximal threat cues, conditional on an imminent attack, and the probability of the observed proximal cues conditional on no imminent attack. Applying Bayes Theorem in odds-ratio form, the posterior odds of an attack are the product of the prior odds of an attack (based on distal threat cues) and the likelihood ratio (or Bayes Factor), equal to the ratio of the conditional probabilities of the observed proximal threat cues, as stated in Equation (1) (Coombs et al. 1970; Swets and Pickett 1982; McFall and Treat 1999; Wickens 2001).

$$\frac{P(H_a \mid T)}{P(H_{\sim a} \mid T)} = \frac{P(T \mid H_a)}{P(T \mid H_{\sim a})} \times \frac{P(H_a)}{P(H_{\sim a})}, \quad (1)$$

where Ha is the hypothesized attack occurs, H~a is the hypothesized attack does not occur, T = the observed behavioral (proximal) threat cues.

For example, for a particular context, the defendant may have a prior probability of an attack of 0.05 (one in twenty, prior odds = 1/19). The proximal threat cues observed may have a probability of 0.20 given that an attack is imminent, while having only probability of 0.10 given that an attack is not imminent. This results in a likelihood ratio of 2 (0.20/0.10). The posterior odds of an attack for this situation would be (1/19)* 2 ~ 0.105, corresponding to a posterior probability of an imminent attack of ~ 0.12.

As indicated in Fig. 4, the decision to act in self-defense depends on the proximal threat cues (likelihood ratio for the observation) and the threshold for taking action. This formulation assumes that the defendant is acting to maximize their expected utility. The threshold for taking action depends on both the prior odds of an attack and the relative costs of a false positive (taking action when there is no actual threat) or a false negative (not taking action when there is a genuine threat). The prior odds of an attack are often quite low based on purely base-rate information alone. Likewise, for the defendant, the costs of a false negative are often relatively high compared to the costs of a false positive, as a false negative may result in severe personal injury or loss of life. Following the previous convention and without loss of generality, we set the utility of correct classifications to zero, reflecting no cost. Hence, the (dis)utility of errors will be negatively valued to reflect the costs of each type of error. The threshold

for proactive self-defensive action is provided in Equation (2) as the product of the prior odds against an attack and the ratio of the cost of a false positive to the cost of a false negative (Coombs, Dawes, and Tversky 1970; Swets and Pickett 1982; McFall and Treat 1999; Wickens 2001; Scurich and John 2010).

$$\frac{P(T \mid H_a)}{P(T \mid H_{\sim a})} > \frac{P(H_{\sim a})}{P(H_a)} \times \frac{U_{tn} - U_{fp}}{U_{tp} - U_{fn}} \quad (2)$$

Where, Utn = utility of a true negative (=0), that is, not acting in self-defense when an attack is not imminent; Utp = utility of a true negative (=0), that is, acting in self-defense when an attack is imminent; Ufp = (dis)utility (cost) of a false positive, that is, acting in self-defense when no attack is imminent; and Ufn = (dis)utility (cost) of a false negative, that is, not acting in self-defense when an attack is imminent.

The value of the right side of the inequality in Equation (2) defines the action threshold. For example, using the same prior odds as above (19:1 against an imminent attack), if the defendant's cost of a false negative was 10 times the cost of a false positive, the action threshold likelihood ratio would be 19/10 = 1.9. For this particular case, the defendant would maximize the expected value (minimize expected cost) by taking proactive defensive action whenever the likelihood ratio (probability of the observed proximal threat cues, conditional on an imminent attack, divided by the probability of the observed proximal threat cues, conditional on no attack) exceeded 1.9. For the proximal threat cues example described above, the likelihood ratio exceeds the normative threshold, 2.0 > 1.9, and the expected utility is maximized (expected cost minimized) by acting in self-defensive.

In the above examples, the criteria for the necessity of self-defensive action are met based on the requirement to maximize expected utility (equivalent to minimizing expected cost). However, the posterior probability of an imminent attack is only 0.12, indicating that, it is more likely than not, an attack is not imminent based on the observed proximal threat cues and prior probability of an attack. Hence, in this case, an act of self-defense should not be deemed justifiable since both criteria are required for justifiable self-defense.

1.3 The current study

While a legally justifiable act of self-defense has a three-pronged requirement, the current paper focuses exclusively on the assessment of imminence—that is, how individuals perceive threat cues in the context of self-defense. Our benchmarks of interest include the base rate of criminal activity based on distal environmental cues (i.e., neighborhood crime rate) and the more proximal information that reveals the threat from the possible attacker (i.e., potential attacker's behavior). Thus, this article examines the relationship between the prior probability of an attack, proximal threat cues, the posterior probability of an attack, and the intent to engage in self-protective behavior.

2. Method

2.1 Design overview

Respondents were recruited from Amazon Mechanical Turk for a vignette study. Each respondent read four threat vignettes involving situations where respondents are approached by threatening persons. We manipulated prior information (a “high crime” versus a “low crime” neighborhood) and threat cues (“suspicious” versus “unsuspicious” behaviors by the possible attacker) counterbalanced across the four scenarios. We measured respondents' perception of threat uncertainty and their binary pre-emptive self-defense judgment. The study concluded with one attention check question, asking respondents about the scenarios they had previously read. Data collection was performed and completed during the summer of 2019.

2.2 Participants

A total of 250 US respondents were recruited on mTurk. By completing this study, each respondent received 0.75 USD. Fifteen respondents were removed for either failing the attention check question or indicating that they wished to have their data excluded. The final sample size is 235, composed of 47.7% (*N* = 112) males and 77% (*N* = 181) non-Hispanic Whites, with a median age of 34 years. Approximately 38.7% (*N* = 91) of the respondents self-identified as “Very Liberal” or “Liberal,” 21.7% (*N* = 51) self-identified as either “Very Conservative” or “Conservative,” and the remaining 39.6% (*N* = 35) self-identified as “Neutral,” “Slightly Liberal,” or “Slightly Conservative.”

Table 1.2 ×2 within-subjects factorial design.

Proximal threat (Bayes factor)	Crime rate (prior odds)	
	High	Level
High	High crime rate	Low crime rate
	High threat cue	High threat cue
Low	High crime rate	Low crime rate
	Low threat cue	Low threat cue

2.3 Procedure

After reading a brief information sheet about the study and obtaining informed consent, respondents sequentially read the four threat scenarios, each containing a situation where a stranger is approaching the respondent. For each of the four threat scenarios, we manipulated (a) distal threat cues, defined as *high vs. low base* rate of criminal activity and (b) proximal threat cues as suspiciousness of the stranger's behavior, creating a 2 (high versus low base rate of danger) × 2 (high versus low likelihood of harm) within-subjects design, displayed in Table 1. Base rates were manipulated for each of the four vignettes (described below) to fit either a high or low risk for crime prior to any interaction with the stranger. Likewise, the proximal threat cues were manipulated in terms of the level of suspiciousness of the stranger's behavior. The four threat cells in Table 1 were paired evenly with the four vignettes described below, such that all four vignettes were used for each of the four cells in Table 1. In addition, the order of presentation of the four vignettes was counterbalanced. This design balances out any vignette or order effects across the four cells of the 2 × 2 design in Table 1.

2.4 Scenario selection

Respondents were asked to imagine themselves in four different scenarios: (1) being followed in a parking garage; (2) withdrawing cash at an ATM while a stranger is advancing; (3) a person approaching on foot in a quiet neighborhood at night; and (4) being followed to the hotel room. Table 2 presents scenario details for each vignette,

including the mode of self-defense, the low and high threat distal cues, and the low and high proximal threat cues (see Appendix A for the four complete vignettes). Each scenario offers a unique location setting and self-defense mechanism. We chose these four scenarios to provide a range of different settings and defense methods to observe whether our manipulations generalize across various realistic contexts.

A Latin Square Design was used to counterbalance the manipulated distal and proximal threat cues to the four scenarios. Respondents were randomly partitioned into four groups, and the cues' distal and proximal threat levels were assigned to scenarios as indicated in Table 3. The 4-group Latin Square Design results in the four combinations of distal and proximal threat cues paired with each of the scenarios only once and appearing in each of the four possible serial positions only once. Thus, the 2 × 2 design described in Table 1 is completely counterbalanced with scenario content and serial order of presentation. After reading each scenario, respondents indicated their perceived threat of an imminent attack from the stranger and indicated whether they would engage in proactive self-defense. The order of these two questions was randomized. Likewise, the order of presenting the two possible mutually exclusive hypotheses about the stranger was also randomized. Finally, an attention check question was administered to ensure data quality, and respondents provided demographic information, including gender, age, race, and political orientation.

2.5 Response measures

Following each scenario, respondents were asked to provide their perceived threat of attack from the stranger in the form of the probabilities that the suspicious person intended versus did not intend harm. Respondents were required to make the two probabilities sum to 1.0. Additionally, they were also asked to indicate their intent to engage in self-defense should they

Table 2. Scenario vignettes: setting and premise, mode of self-defense, distal context manipulations, and proximal threat manipulations.

	Scenario A	Scenario B	Scenario C	Scenario D
Premise	Followed in a parking garage	Approached from behind at ATM at night	Approached in neighborhood at night	Followed to hotel at night
Self-defense method:	Pepper spray	Shove to ground	Strike with steel object	Throw small rock
Prior context information manipulation	Low/High crime neighborhood	Low/High crime neighborhood	Low/High crime neighborhood	Low/High crime neighborhood
Proximal threat cues manipulation	Other cars in the garage (yes/no)	Other ATM's available (no/yes)	Other houses nearby that person could be walking to (yes/no)	Resembles someone who followed you earlier (no/yes)

Table 3. Latin square design for assigning levels of distal and proximal threat cues to scenarios

	Group 1	Group 2	Group 3	Group 4

Scenario A	High crime rate	Low crime rate	Low crime rate	High crime rate
(Parking garage)	Low threat cue	High threat cue	Low threat cue	High threat cue
Scenario B	High crime rate	High crime rate	Low crime rate	Low crime rate
(ATM)	High threat cue	Low threat cue	High threat cue	Low threat cue
Scenario C	Low crime rate	High crime rate	High crime rate	Low crime rate
(Neighborhood)	Low threat cue	High threat cue	Low threat cue	High threat cue
Scenario D	Low crime rate	Low crime rate	High crime rate	High crime rate
(Hotel)	High threat cue	Low threat cue	High threat cue	Low threat cue

find themselves in the situation described. The intent was a dichotomous response of either engaging or not engaging in self-defense at the time corresponding to the end of the vignette. The order of presentation of the two response measures, perceived threat and intent to engage in self-defense, was randomized.

3. Results

3.1 Analysis overview

Figure 5 presents a conceptual model, including the manipulated distal and proximal threat cues, and dependent measures of the perceived threat and intent to engage in self-defense. First, we report results on the independent and interactive effects of the distal and proximal threat cues on the perceived threat, accounting for respondent demographic variables (gender, age, race, and political orientation). Second, we examine the relationship between threat perception and intent to engage in self-defense for all four combinations of levels of distal and proximal threat cues, accounting for respondent demographic variables. Finally, we focus on estimating the threat perception threshold for acting in self-defense. In all analyses, age is measured in years, gender (female versus male) and race (non-Hispanic White versus Other) are dichotomous, and political orientation is a categorical variable (liberal, moderate, conservative).

3.2 Perceived threat

Perceived threat appraisal was assessed as the respondent's indicated probability of an imminent attack on a 0–100% scale. We conducted a 2 (low crime rate, high crime rate) × 2 (low proximal threat, high proximal threat) × 2 (male, female respondent gender) × 3 (liberal, moderate, conservative respondent political orientation) mixed-model Analysis of Covariance (ANCOVA) on the perceived threat. The first two factors are manipulated within-subjects factors, while the

Figure 5. Conceptual model of manipulated factors (rounded rectangles) and dependent measures of threat and intent to engage in self-defense (ellipses).

last two are individual difference variables hypothesized as relevant for threat perception. A total of four covariates were included in the model to account for variation due to respondent age (years, centered at 40), race/ethnicity

(white not-Hispanic versus all others), order of presentation of the questions for the two dependent variables (threat perception probability before self-defense intent versus self-defense intent before threat perception judgment), and order of presentation of the two response options for self-defense intent (act before not act versus not act before act). The mean perceived threat of imminent attack ranged from 0.40 (low crime rate, low proximal threat behavior) to 0.54 (high crime rate, high proximal threat behavior).

Estimated marginal perceived threat probabilities from the ANCOVA model, ranging from 0.41 to 0.56, are plotted in Fig. 6 for the two within-subject manipulated variables. Both the distal cue manipulation (crime base rate) and the proximal threat manipulation (stranger's behavior) had the predicted impact on respondents' threat probabilities. High crime base rates increased threat probabilities by about 0.07 over low crime base rates [$F(1,224) = 11.37$, $P = 0.001$, partial $\eta^2 = 0.05$]. Likewise, highly suspicious behavior exhibited by the stranger increased threat probabilities by about 0.09 over moderately suspicious behavior [$F(1,224) = 31.13$, $P < 0.001$, partial $\eta^2 = 0.12$]. There was no significant interaction ($F < 1.0$) between the two within-subjects manipulated factors. Notably, none of the four covariates significantly predicted threat perceptions.

The ANCOVA indicated that respondents' political orientation influenced the threat appraisals and moderated the impact of distal threat cues (crime rate) on threat appraisals. As shown in Fig. 7, threat appraisals were overall higher for conservatives ($M = 0.55$) compared to moderates ($M = 0.47$) and liberals ($M = 0.43$) [$F(2,224) = 4.57$, $P = 0.011$, partial $\eta^2 = 0.04$]. The ANCOVA also indicated that the impact of distal threat cues (crime rates) on overall threat perceptions depended on the respondent's political orientation. As indicated in Fig. 7, threat perception probabilities were about 0.06 higher in the high crime compared to the low crime scenarios for political moderates. However, this difference in threat perception probability was attenuated by two-thirds (to about 0.02) for political liberals and magnified to nearly double (about 0.11) for political conservatives [$F(2,224) = 3.55$, $P = 0.03$, partial $\eta^2 = 0.03$]. This result suggests that liberals essentially ignore distal cues such as crime rate in formulating threat perceptions. At the same time, conservatives weigh distal cues such as crime rate heavily in appraising threats of an imminent attack.

Regardless of the distal and proximal threat cues, political orientation moderates the differences in threat appraisal for males and females. As indicated in Fig. 8, non-conservative females report greater imminent threat than males (0.10 and 0.14, respectively, for liberals and moderates.) However, conservative male appraisals of the threat of imminent attack are greater than conservative female appraisals by about 0.10. While female threat perception probabilities are about the same across all political orientations, male threat perception probabilities depend

Figure 6. Estimated marginal threat probabilities of an imminent attack by distal and proximal threat cues.

Figure 7. Estimated marginal threat probabilities of an imminent attack by distal threat cues and political orientation.

on political orientation, ranging from 0.38 (liberals) and 0.40 (moderates) to 0.60 (conservatives). The cross-over interaction between political orientation and gender is significant [F (2,224) = 5.57, $P = 0.004$, partial $\eta^2 = 0.05$].

Figure 8. Estimated marginal threat probabilities of an imminent attack by political orientation and gender.

3.3 Intention to act in self-defense

The observed proportion of respondents indicating an intent to act in self-defense is presented in Fig. 9 for each of the four combinations of crime rate and behavioral intensity of the threat. Across all four combinations of threat cues, barely 30% indicated an intention to engage in self-defense. Notably, the combination of high proximal threat

behavior in a high crime rate area resulted in a significantly higher percentage (42%) of individuals who intend to act in self-defense compared to the other three threat scenarios. Highly threatening behavior is more likely to result in proactive self-defense in a high-crime area than in a low-crime area.

3.4 Threshold of likelihood of imminent attack for intent to act in self-defense

We conducted binary logistic regression (BLR) analyses on the dichotomous decision to act in self-defense for each of the four threat scenarios. Predictors included the same four individual difference variables and the two dichotomous order variables (all contrast coded) included in the analyses on perceived threat probabilities reported above. None of these BLR models was significantly better than predicting the self-defense decision using the overall proportions plotted in Fig. 9 for each scenario [χ^2 , df(7) = 12.12, 10.59, 10.04, and 7.04; all *P*'s> 0.05]. Knowledge of gender, age, race/ethnicity, and political orientation did not improve the prediction of self-defense intent over simply knowing the observed proportions intending self-defense.

We repeated the BLR analyses reported above and included the reported perception of imminent threat of attack as a seventh predictor, continuing to account for the individual difference variables and the task ordering variables. Parameter estimates are presented for each of the four models in Table 4. All four models significantly improved prediction of engaging in self-defense over the null model, as indicated by the significant Chi-Square statistics reported at the bottom of Table 4. The respondents' perceived probability of imminent attack was a significant predictor ($P < 0.001$ in all four models). Holding all individual difference and order predictors constant at the centered value of 0.0, an increase of 0.1 (10%) in the perceived probability of imminent attack, e.g. from $P = 0.25$ to $P = 0.35$, increased odds of acting in self-defense by an odds ratio [OR = exp(B)$^{\text{Unit-change}}$] between 2.2 (= 1.082^{10}) and 2.5 (= 1.096^{10}), depending on the threat scenario.

As indicated in Table 4, various individual differences and order effects were also observed. Non-whites, compared to whites, were more likely to indicate an intent to engage in

Figure 9. Observed proportion of respondents with intent to act in self-defense by distal (crime rate) and proximal (behavioral) threat cues.

Table 4. BLR for predicting intent to act in self-defense (1 = defend self; 0 = no self-defense).

Distal prior knowledge	**High crime rate**		**High crime rate**		**Low crime rate**		**Low crime rate**	
Proximal threat	**High threat**		**Low threat**		**High threat**		**Low threat**	
Proximal threat signal								
Predictor variable	**OR = exp(B)**	**SE (B)**	**OR = exp(B)**	**SE (B) O**	**OR = exp(B)**	**SE (B)**	**OR = exp(B)**	**SE (B)**
Constant	0.007	0.672	0.005	0.730	0.003	0.775	0.002	0.794
Age (years)	0.992	0.018	0.974	0.020	0.992	0.018	0.972	0.022
Gender (Male	1.366	0.373	0.868	0.404	0.519	0.421	1.267	0.450

versus Female)								
Political orient. (Linear)	0.885	0.359	0.404**	0.410	0.999	0.402	0.542	0.434
Political orient. (Quad.)	0.936	0.306	0.786	0.334	0.779**	0.337	0.740	0.374
Ethnicity (W versus Other)	0.908	0.441	2.497*	0.457	2.725*	0.468	0.878	0.491
Question order	0.446*	0.389	0.541	0.400	0.509	0.409	0.335*	0.462
Response order	1.277	0.376	0.658	0.398	1.384	0.420	0.827	0.441
P (imminent attack) in %	1.082***	0.010	1.083***	0.011	1.092***	0.012	1.096***	0.013
Chi-Square, df = 8	130.79***		103.33***		132.34***		111.34***	

** <0.05;*

*** <0.01;*

*** <0.001.*

Note. Gender = Male (−0.5) versus Female (0.5); Ethnicity = Non-Hispanic White (−0.5) versus Others (0.5); Political Orientation Linear = Liberal (−0.707) versus Conservative (0.707); Political Orientation Quadratic = Moderate (−0.816) versus Others (0.408); Question order = Judgement to defend followed by Probability of attacker intending harm (−0.5) versus Probability of attacker intending harm followed by Judgment to defend (0.5); Response Order = Act in self-defense followed by Refrain from acting (−0.5) versus Refrain from acting followed by Act in self-defense (0.5).

self-defense, given a reported probability of imminent attack, in both the High–Low (OR = 2.497) and Low–High (OR = 2.725) scenarios. Liberals were more likely to indicate an intent to engage in self-defense than conservatives in the High–Low (OR = $0.404^{-1.414} = 3.60$)

scenario, conditional on a given probability of imminent attack. Finally, judging intent to act in self-defense before estimating the threat of imminent attack increased the odds of self-defense intent in the High–High (OR = $0.446^{-1} = 2.24$) and Low–Low (OR = $0.335^{-1} = 2.99$) scenarios. (We follow the common practice of reporting the reciprocal of odds ratios less than 1.0 and stating the finding by reversing the contrast coding.)

The BLR parameter estimates were used to construct plots for each threat scenario predicting the respondent's

probability of acting in self-defense as a function of the respondent's perceived threat of imminent attack. The plots, displayed in Fig. 10, were calculated using a value of 0 for the contrast-coded individual difference and task order variables that were contrast-coded for the analyses. These curves average overall individual difference and order covariates and do not correspond to estimates of any particular individual (e.g. a 32-year-old Hispanic liberal female). Instead, these curves reflect expectations for a randomly sampled individual.

These four curves all have similar shapes, and all converge on approximately the same perceived threat threshold for acting in self-defense. The curves cross the p(intent to act in self-defense) = 0.50 horizontal line at values of perceived threat between 0.62 and 0.65. This result suggests that respondents perceived threat of an imminent attack judgments account for all information about the threat situation, including the distal crime rate threat cues and the proximal behavioral threat cues that were manipulated. Respondents' decision to act in self-defense appears to be determined by their perception of a threat of imminent attack. The empirical thresholds indicated in Fig. 10 suggest a threshold of greater than 60% probability of an imminent attack for respondents to be equally likely to act or not in self-defense.

4. Discussion

The current study investigated how individuals assess threat cues in the context of self-defense. We manipulated distal threat cues (high-versus low-crime neighborhood) and proximal threat cues (high versus low likelihood of harm) using a 2 × 2 within-subjects design. US adult respondents read four self-defense scenario vignettes, in which distal and proximal threat cues counterbalanced across the scenarios. The respondents then reported their perception of threat uncertainty and their binary pre-emptive self-defense judgment.

Both distal threat cues and proximal threat cues influenced respondents' perceived probability of imminent harm. Respondents reported a higher probability of harm for high-threat observations and in high-crime rate contexts. Remarkably, respondents utilized the base rate of crime in appraising the threat of an imminent attack and combined this information with behavioral observations of the threatening individual with no indication of a synergistic (interactive) effect. These findings are consistent with PMT (Rogers 1975, 1983) and PADM (Lindell and Perry 2004, 2012), showing that both a person's prior beliefs about adverse events and the assessment of immediate threat cues can affect their perception of threat and intent to act in self-defense.

Respondents' political orientation influenced their threat appraisals. For example, conservative respondents weighed distal cues significantly heavier than liberal respondents when formulating threat perceptions. Political orientation also significantly moderated the differences in threat appraisal between males and females. While female respondents' perceived probability of harm remained about the same across all political orientations, male threat perception probabilities depended on their political orientation; conservative male respondents reported a significantly higher perceived likelihood of harm compared to moderate and liberal male respondents.

Across all four scenarios, intentions to act in self-defense were strongly associated with the perceived threat, assessed as the probability of an imminent attack. While respondents' political orientation and gender impacted their appraisal of the threat of an imminent attack, the decision to act in self-defense could not be predicted by respondents' demographics only. The respondents' decision to act in self-defense corresponds with the legal requirements for justifiable self-defense, which requires a belief of imminent threat of harm and a need to act to counter the threat.

Logistic regression analysis was used to estimate (even odds) thresholds, ranging from 0.62 to 0.65, to act in self-defense as a function of the perceived threat probability for each of the four scenarios. Empirical threat probability thresholds, between 0.62 and 0.65, to engage in

Figure 10. Estimated probability of acting in self-defense conditional on perception of threat of imminent attack, crime rate, and intensity of threatening behavior.

self-defense indicate that the odds of an imminent attack must be nearly 2:1 to trigger proactive self-defense in each of the four urban threat scenarios. Thus, the threshold probability of imminent threat required for intent to act in self-defense is greater than a standard of “more likely than not” to attack (>50% by definition) but less than “beyond a reasonable doubt (BRD)” of an imminent attack (usually >90 or 95%). Remarkably, all four threshold estimates were approximately the same across four distal and proximal threat level combinations. This finding suggests that particular distal cues (high or low crime rate) and particular proximal cues (observed behavior) did not affect the threshold in perceived threat for acting in self-defense after accounting for respondents' appraisal of attack imminence. That is, all information in the manipulated factors (distal prior knowledge and proximal threat signals) were captured in the respondents' probability judgment of an imminent attack.

We want to clarify that we do not suggest that individuals conduct a subjective Bayesian calculation or formally calculate costs and make choices to minimize expected costs. Instead, we compare our empirical results to these formal models to derive insights into the implied costs of mistaken self-defense and failure to take action to thwart an actual threat. Failing to engage in proactive self-defense involves the possibility of property loss, injury, or even death. Mistaken self-defense involves all of these possible outcomes and may provoke an attack or increase the severity of a planned attack if the defensive action is not effective. Moreover, subsequent mistaken self-defense costs could also include civil liability and criminal culpability.

5. Limitations

As with any vignette study, there are several limitations to the current experiment. All scenarios used for this study were constructed in an isolated urban environment during the nighttime.

While we incorporated different scenarios where one may employ self-defense, we recognize that these scenarios do not capture all of the nuances of threat situations (e.g. other times of day, rural and suburban areas, the presence of law enforcement and bystanders, etc.). Future research should explore these situational factors and the extent to which our findings generalize to these other contexts.

Another important caveat is that all scenarios were simulated situations that did not include emotional responses from “in-the-moment” experiences. Our empirical study likely induces more System 2 processing than expected in an actual threat context “on the street.” Nevertheless, we believe that the scenarios employed were highly realistic and relatable, based on comments from respondents. No one can predict precisely how they will respond in a threat context until they are in the situation. However, respondents' threat perception and intention to act in self-defense reflect both threat appraisals and decisions to take action that are indicative of how people would expect to respond in these highly relatable urban situations.

Acknowledgements

We thank Dr Matthew Baucum for his assistance in conducting this research.

Funding

The authors declare that they have no known conflicts of interest associated with this study. This material is based upon work supported by the U.S. Department of Homeland Security under [Grant Award 22STESE00001-02-00]. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland

Security. This award was made to Northeastern University and the University of Southern California is a subawardee.

Appendix A: Sample scenarios

Scenario: Followed in a parking garage

After having some dinner with friends at a downtown restaurant, you decide to head home at around 11 p.m. You parked in a parking garage that none of your friends ended up using, so you walk back to your car alone, keeping in mind that the neighborhood you are in happens to have fairly **high/low crime**. You take the elevator to the third floor, exit the elevator, and begin walking toward your car on the other side of the garage. As you begin walking, you notice, out of the corner of your eye, someone beginning to walk in your direction as you exit the elevator. You continue walking toward your car and, as you do, you hear footsteps following you. They persist for about 20 s as you continue making your way across the parking garage floor. You see no one else in the parking garage, and upon looking around, you see **no other cars/a few other cars** besides yours in the immediate vicinity. You decide to speed up your walking pace, and you think you hear the steps quicken as well. You remember that you have a small can of pepper spray in your laptop bag, and know that you are probably within the range to be able to spray the person behind you and give you enough time to get away in your car. You consider turning around and using the pepper spray before the person has a chance to get any closer.

Scenario: Approached from behind at ATM at night

You are leaving work late one night and decide to take a taxi home. Realizing that you do not have any cash, you decide to stop at an ATM to pull some out. As you approach the ATM, you notice that the street is empty except for one person standing across the street from you. You begin using the ATM and glance behind you to see that the person who was standing across the street is now crossing toward you. You know you are in a fairly **unsafe/safe part of town**; just to be safe, you try to hurry up and leave the ATM as fast as possible. You hear footsteps approaching as your transaction is still being processed, and realize the person has now crossed to your side of the street and is waiting behind you. You look and notice that this is **the only ATM**

available for use/two other ATMs next to this one. As your transaction is finishing, you look to your left and see a taxi cab, waiting about a block away, that you could quickly get to if you had a little bit of time. You think that if you were to turn around and push the person behind you to the ground, you would be able to make it to the cab with enough time to flee. At this point, the ATM begins distributing your cash, and you hear the person behind you take a step forward.

Scenario: Approached in neighborhood at night

It is a Saturday night, and you have been invited to go spend some time at a friend's house, who lives about 15 min away. You agree to go and get into your car to drive over. Since it is already late (around 10 p.m.), there are not many other cars on the road. About 10 min into the drive, you notice another car behind you on the road. Over the course of the next 5 min, you make about three different turns to arrive at your friend's house, all of which the car behind you also makes. When you arrive at your friend's neighborhood, the street is fairly full with parked cars, but you see a parking spot about a block away from your friend's house. You look behind you to see the car that was following you is still behind you—you know that your friend lives in a relatively **high-/low-crime neighborhood**, so just to be safe, you slow down near the open parking spot to see if the car behind you stops as well. At that point, the car drives around you and proceeds down the street, and with it no longer behind you, you figure it is safe to go ahead and park. After pulling into the parking spot, you decide to store some of your things in your trunk so that

they do not remain exposed in the car. You exit the car and open your trunk, at which point you notice that the car that had been following you has parked just a few cars down from yours. You see that the driver has just exited the car and is quickly walking down the sidewalk toward you, wearing a pullover sweatshirt and carrying something small in their right hand. You glance around and notice that there are **no apartments or houses in the immediate vicinity/ a few apartments and houses in the immediate vicinity, which the person could be walking to**. The person is now about 10 feet away, and, realizing that you do not have enough time to get back in your car, you quickly look down and see a tire iron (a long, curved metal object used for fixing tires) sitting in your trunk. You consider using it to strike the driver of the other car as they approach to give you enough time to enter your car and drive away.

Scenario: Followed to a hotel at night

You are traveling one weekend and are staying in a hotel near the downtown area of the city you are visiting. After going to dinner and spending some time downtown, it is getting late and you decide to return to your hotel for the night. You decide to walk back since it is only about six blocks away, keeping in mind that the downtown area of this particular city has seen **a recent increase/decrease in crime**. About one block into your walk, you take a right turn and accidentally bump into someone who was standing on the sidewalk where you were turning. You apologize and keep walking, and as you do you hear footsteps behind you. You glance back to see that the person you bumped into is now following about 15 feet behind you—they seem to be wearing shorts and a t-shirt with a jacket on and are walking fairly briskly. This continues for about two blocks, at which point the person behind you has made two of the same turns you have. You decide to run the rest of the distance to the hotel in order to get back as fast as possible; you look behind you a few seconds later to see that the person who was following you is no longer there.

About 2 min later, you arrive at the entrance to your hotel. You begin looking for your key card in your wallet (since the doors are locked during the night), and as you are searching for it, you hear footsteps around the corner. You look up to see someone turning the corner and running very quickly toward you—they are difficult to make out, but seem to be wearing shorts and a t-shirt, **but no jacket/and a jacket**. You quickly grab your key card, but realize that you do not have enough time to enter the hotel before the person reaches you. You look down and see a small rock sitting on the ledge next to you; you think that if the person actually intends to harm you, you could throw the rock at them and buy just enough time to run around to another hotel entrance.

References

American Law Institute. (1962) *Model penal code*. Philadelphia, PA: The American Law Institute.

BAR-HILLEL, M. (1980) 'The Base-rate Fallacy in Probability Judgments', *Acta Psychologica*, **44**: 211–33. http://dx.doi.org.libproxy1.usc.edu/10.1016/0001-6918(80)90046-3

BETSCH, T., et al. (1998) 'Natural Sampling and Base-rate Neglect', *European Journal of Social Psychology*, **28**: 269–73.

CHRISTENSEN-SZALANSKI, J. J., and BEACH, L. R. (1982) 'Experience and the Base-rate Fallacy', *Organizational Behavior and Human Performance*, **29**: 270–78.

COHEN, N. P., JOHNSON, M. G., and HENLEY, T. B. (1993) 'The Prevalence and Use of Criminal Defenses: A Preliminary Study', *Tennessee Law Review*, **60**: 957–82.

COOMBS, C. H., DAWES, R. M., and TVERSKY, A. (1970) *Mathematical Psychology: An Elementary Introduction*. Englewood Cliffs, NJ: Prentice-Hall.

FEDERAL BUREAU OF INVESTIGATION. (2020) *National Incident-Based Reporting System Report*. Washington, DC: U.S. Department of Justice.

FERZAN, K. K. (2005) 'Justifying Self-defense', *Law and Philosophy*, **24**: 711–49.

FLETCHER, G. P. (1996) 'Domination in the Theory of Justification and Excuse', *University of Pittsburgh Law Review*, **57**: 553–78.

FLOYD, D. L., PRENTICE-DUNN, S., and ROGERS, R. W. (2000) 'A Meta-analysis of Research on Protection Motivation Theory', *Journal of Applied Social Psychology*, **30**: 407–29.

FONTAINE, R. G. (2010) 'An Attack on Self-defense', *American Criminal Law Review*, **47**: 57–90.

GOODIE, A. S., and FANTINO, E. (1999) 'What Does and Does Not Alleviate Base-rate Neglect Under Direct Experience', *Journal of Behavioral Decision Making*, **12**: 307–35.

HEATH, R. L., et al. (2018) 'Risk Communication Emergency Response Preparedness: Contextual Assessment of the Protective Action Decision Model', *Risk Analysis*, **38**: 333–44.

HELLER, K. J. (1998) 'Beyond the Reasonable Man: A Sympathetic but Critical Assessment of the Use of Subjective Standards of Reasonableness in Self-defense and Provocation Cases', *American Journal of Criminal Law*, **26**: 1–120.

KIM, J. Y. (2008) 'The Rhetoric of Self-defense', *Berkeley Journal of Criminal Law*, **13**: 261–305.

KOEHLER, J. J. (1996) 'The Base Rate Fallacy Reconsidered: Descriptive, Normative, and Methodological Challenges', *Behavioral and Brain Sciences*, **19**: 1–17

KRYNSKI, T. R., and TENENBAUM, J. B. (2007) 'The Role of Causality in Judgment Under Uncertainty', *Journal of Experimental Psychology: General*, **136**: 430–50. http://dx.doi.org.libproxy1.usc.edu/10.1037/0096-3445.136.3.430

LINDELL, M. K., and PERRY, R. W. (2004) *Communicating Environmental Risk in Multiethnic Communities*. Thousand Oaks CA: Sage.

LINDELL, M. K., and PERRY, R. W. (2012) 'The Protective Action Decision Model: Theoretical Modifications and Additional Evidence', *Risk Analysis*, **32**: 616–32.

LOVETT, M. C., and SCHUNN, C. D. (1999) 'Task Representations, Strategy Variability, and Base-rate Neglect', *Journal of Experimental Psychology: General*, **128**: 107.

LYON, D., and SLOVIC, P. (1976) 'Dominance of Accuracy Information and Neglect of Base Rates in Probability Estimation', *Acta Psychologica*, **40**: 286–98. https://doi.org/10.1016/0001-6918(76)90032-9

MADDUX, J. E., and ROGERS, R. W. (1983) 'Protection Motivation and Self-efficacy: A Revised Theory of Fear Appeals and Attitude Change', *Journal of Experimental Social Psychology*, **19**: 469–79.

MCFALL, R. M., and TREAT, T. A. (1999) 'Quantifying the Information Value of Clinical Assessments with Signal Detection Theory', *Annual Review of Psychology*, **50**: 215–41.

MCMAHAN, J. (1994) 'Self-defense and the Problem of the Innocent Attacker', *Ethics*, **104**: 252–90.

MILNE, S., SHEERAN, P., and ORBELL, S. (2000) 'Prediction and Intervention in Health-related Behavior: A Meta-analytic Review of Protection Motivation Theory', *Journal of Applied Social Psychology*, **30**: 106–43.

NEWMAN, I. R., GIBB, M., and THOMPSON, V. A. (2017) 'Rule-based Reasoning is Fast and Belief-based Reasoning can be Slow: Challenging Current Explanations of Belief-bias and Base-rate Neglect', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **43**: 1154.

NOURSE, V. F. (2001) 'Self-defense and Subjectivity', *The University of Chicago Law Review*, **68**: 1235–308.

OLESON, K. C., and DARLEY, J. M. (1999) 'Community Perceptions of Allowable Counterforce in Self-defense and Defense of Property', *Law and Human Behavior*, **23**: 629–51.

PILDITCH, T. D., LAGATOR, S., and LAGNADO, D. (2021) 'Strange but True: Corroboration and Base Rate Neglect', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **47**: 11.

ROGERS, R. W. (1975) 'A Protection Motivation Theory of Fear Appeals and Attitude Change', *The Journal of Psychology*, **91**: 93–114.

ROGERS R. W. (1983) 'Cognitive and Physiological Processes in Fear-based Attitude Change: A Revised Theory of Protection Motivation'. In: Caccioppo, J., and Petty, R., (eds) *Social Psychophysiology: A Sourcebook*. New York: Guilford.

ROGERS, R. W. (1985) 'Attitude Change and Information Integration in Fear Appeals', *Psychological Reports*, **56**: 179–82.

SCURICH, N., and JOHN, R. (2010) 'The Normative Threshold for Psychiatric Civil Commitment', *Jurimetrics: The Journal of Law, Science & Technology*, **50**: 425–52.

SLOMAN, S. A. (2003) 'Causal Models can be Used to Predict Base-Rate Neglect', *Proceedings of the Annual Meeting of the Cognitive Science Society*, **25**: 1088–1093. Retrieved 7 July 2022, from https://escholarship.org/uc/item/1px5b0rw

SWETS, J., and PICKETT, R. (1982) *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory*. New York: Academic Press.

TURPIN, M. H., et al. (2020) 'The Environmental Malleability of Base-rate Neglect', *Psychonomic Bulletin & Review*, **27**: 385–91.

TVERSKY, A., and KAHNEMAN, D. (1974) 'Judgment under Uncertainty: Heuristics and Biases', *Science*, **185**: 1124–31.

VALDES, S. G. (2005) 'Frequency and Success: An Empirical Study of Criminal Law Defenses, Federal Constitutional Evidentiary Claims, and Plea Negotiations', *University of Pennsylvania Law Review*, **153**: 1709–814.

WALLERSTEIN, S. (2005) 'Justifying the Right to Self-defense: A Theory of Forced Consequences', *Virginia Law Review*, **91**: 999–1035.

WANG, T. R., et al. (2020) 'Identification of Protective Actions to Reduce the Vulnerability of Safety-critical Systems to Malevolent Intentional Acts: An Optimization-based Decision-making Approach', *Risk Analysis*, **40**: 565–87.

WELSH, M. B., and NAVARRO, D. J. (2012) 'Seeing is Believing: Priors, Trust, and Base Rate Neglect', *Organizational Behavior and Human Decision Processes*, **119**: 1–14.

WICKENS, T. D. (2001) *Elementary Signal Detection Theory*. Oxford: Oxford University Press.

End of Document