Behavior Research Methods
https://doi.org/10.3758/s13428-023-02066-9

Characterizing the role of impulsivity in costly, reactive aggression using a novel paradigm

Kimberly L. Meidenbauer[1,2] · Kyoung Whan Choe[1,3] · Akram Bakkour[4,5] · Michael Inzlicht[6] · Michael L. Meidenbauer[1] · Marc G. Berman[1,5]

Accepted: 13 January 2023

Abstract

A lack of self-control has long been theorized to predict an individual's likelihood to engage in antisocial behaviors. However, existing definitions of self-control encompass multiple psychological constructs and lab-based measures of aggression have not allowed for the examination of aggression upon provocation where self-control is needed most. We introduce two versions of a novel paradigm, the Retaliate or Carry-on: Reactive AGgression Experiment (RC-RAGE) to fill this methodological gap. Using large online samples of US adults ($N = 354$ and $N = 366$), we evaluate to what extent dispositional impulsivity, self-control, aggression, and state anger contribute to aggression upon provocation when there is a financial cost involved. Results showed that costly retaliation on this task was related to trait aggression and being in an angry emotional state, but not related to social desirability. Importantly, we show that the tendency to act impulsively is a better predictor of costly retaliation than other forms of self-control, such as the ability to delay gratification, resist temptation, or plan ahead. As a browser-based task, the RC-RAGE provides a tool for the future investigation of reactive aggression in a variety of experimental settings.

Keywords Aggression · Impulsivity · Self-control · Anger · Retaliation · Browser-based task

The relationship between self-control and violence has been observed for decades and generated several theoretical accounts of aggression, beginning with Gottfredson & Hirschi's General Theory of Crime (Gottfredson & Hirschi, 1990) and more recently described by the general aggression model (DeWall et al., 2011) and the I^3 theory (Finkel et al., 2012), among others. Despite dissimilarities in what specific, mechanistic role self-control plays in preventing reactive aggressive behavior, these theories generally agree on a common sequence of events. First, some sort of provocation occurs, which triggers the desire to aggress or retaliate. If the provoked individual has sufficient self-control, they will successfully inhibit this desire, and if their self-control is impaired or insufficient to inhibit the aggressive response, they will retaliate (Denson et al., 2012). This type of impulsive aggression upon provocation is referred to as *reactive aggression*, which is often distinguished from premeditated, *proactive aggression* (Barratt et al., 1999; Walters, 2008; cf. Bushman & Anderson, 2001).

Some of the most compelling evidence for this link comes from a recent meta-analysis of 99 observational studies which demonstrated a robust correlation between self-control and deviant or criminal behavior (Vazsonyi et al., 2017). Importantly, there are key environmental influences which lead to lower self-control and similar psychological processes. Research demonstrates that chronic stress, unemployment, resource scarcity, environmental instability, and other stressors can have a significant influence on self-control (Lovallo, 2013; Sheehy-Skeffington, 2020). Thus, the

✉ Kimberly L. Meidenbauer
k.meidenbauer@wsu.edu

✉ Marc G. Berman
bermanm@uchicago.edu

[1] Department of Psychology, Environmental Neuroscience Lab, The University of Chicago, Chicago, IL, USA

[2] Present Address: Department of Psychology, Washington State University, Pullman, WA, USA

[3] Mansueto Institute for Urban Innovation, The University of Chicago, Chicago, IL, USA

[4] Department of Psychology, Memory and Decision Lab, The University of Chicago, Chicago, IL, USA

[5] The Neuroscience Institute, The University of Chicago, Chicago, IL, USA

[6] Department of Psychology, University of Toronto, Toronto, ON, Canada

Published online: 17 February 2023

relationship between self-control and aggression is likely a complex interaction between dispositional and environmental factors.

The notion of self-control is colloquially defined as "willpower" but is used by researchers to describe a number of psychological processes that allow individuals to regulate behavior. As such, self-control does not have a single agreed upon operational or conceptual definition. In Gottfredson and Hirschi's (1990) original definition, self-control is a trait-level construct associated with characteristics such as the ability to: delay gratification, be persistent, exert caution, and inhibit aggressive responses when frustrated. Other research describes self-control as a conscious effort to control one's behavior in the moment when presented with two competing or conflicting goals, and is therefore treated more as a decision-making process that is influenced by both dispositional and situational/environmental factors (Berkman et al., 2017; Hofmann et al., 2009; Inzlicht et al., 2021; Inzlicht & Schmeichel, 2012). Finally, self-control is also conceptualized as the process of choosing a cognitively demanding, context-dependent mode of responding over a more automatic, habit-based or heuristic mode (Boureau et al., 2015).

However, in the extant literature, poor self-control is also described as high impulsivity or poor self-regulation, despite evidence that these may reflect separable psychological processes (Inzlicht et al., 2021). A recent data-driven factor analysis demonstrated that the higher-order self-control construct could actually be broken into two dominant clusters of related behaviors—one most related to impulsivity, reward sensitivity, goal-directedness and mindfulness, and the other loading onto longer-term attitudes surrounding goals, such as grit or will power (Eisenberg et al., 2019). Indeed, many researchers specifically focus on the link between impulsivity and aggression (Barratt et al., 1999; García-Forero et al., 2009) rather than self-control more broadly.

A predominant neurobiological model of aggression is based on the idea that reactive aggression is more likely to occur when individuals have heightened limbic reactivity to provocation and insufficient inhibitory control from prefrontal cortical (PFC) regions (da Cunha-Bang et al., 2017; Nelson & Trainor, 2007; Siever, 2008). This framework of aggression is also described as reflecting a mismatch between a heightened "drive" and an insufficient "brake" when provocation occurs. Evidence for this comes from observed functional and structural abnormalities of the prefrontal cortex and limbic regions such as the amygdala and anterior cingulate cortex in those with a history of aggressive, antisocial behaviors (Best et al., 2002; Raine, 2008; Siever, 2008). However, it remains unclear what the relative importance of self-control (broadly reflecting delayed gratification, resisting temptation, perseverance, etc.; Eisenberg et al., 2019) and impulsivity (i.e., an impaired drive/brake system) are for preventing an aggressive response upon provocation.

One caveat of the work linking self-control impairments and impulsivity to aggression, crime, and violent behaviors is that it has primarily been conducted using observational studies, rather than empirical tests. As evidence mounts for a robust link between self-control/impulsivity and reactive aggression based on this work, an important next step is to empirically identify the most important trait-level (i.e., self-control, impulsivity) and state-level (i.e., situational/environmental cues, emotional state) predictors. Given the ethical and logistical issues that arise when attempting to utilize a laboratory-based measure of aggression, this is no simple task and existing measures of retaliatory aggression are somewhat limited (Lobbestael, 2015; McCarthy & Elson, 2018; Ritter & Eslea, 2005; Tedeschi & Quigley, 1996). While these paradigms may be effective in many contexts, they are not suited to examine aggression where self-control is needed most: where there is an explicit conflict between a desired response (react aggressively) and the correct response (ignore provocation).

For example, these paradigms often elicit aggressive behavior in a context where there may be either explicit or implicit permissibility and encouragement to act aggressively (i.e., Teacher/Learner paradigm; Buss, 1961). While the often-used Competitive Reaction Time Task (Taylor, 1967) does measure reactive aggression upon provocation, this task is embedded within a competitive context where acting aggressively may, in fact, imbue a tactical advantage (Tedeschi & Quigley, 1996). At a minimum, the Competitive Reaction Time Task creates a context where the desired aggressive response is not discouraged. By imbuing a potential incentive or advantage to aggressing, tasks of this nature are not well suited to studying aggression where self-control or inhibition of an impulsive response is needed, as there is no conflict between what is the "right" choice and what is the "desired" response.

In contrast, other paradigms such as the Point-Subtraction Aggression Paradigm, or PSAP (Cherek et al., 1996), are able to evaluate aggression upon provocation that has a cost involved, but it does not allow for the examination of impulsive aggression. In the typical PSAP and its close variants, participants press a button to gain money and an opponent will occasionally steal some of their earnings. Depending on the specific version used, participants can subtract points from their opponent, ignore their opponent's actions, or protect their money. While this paradigm has been shown to distinguish between participants with and without a history of violence (Cherek et al., 1996; Cherek et al., 2000), it is not ideally suited to study impulsive, reactive aggressive responses as the participant cannot retaliate against their opponent immediately. If participants are provoked while pressing the key used to earn money, they must wait until they've finished that round of key presses before retaliating. Thus, there is a temporal delay between the time that a person experiences provocation and when they are actually able to retaliate. Consistent with this limitation, a study found that participants high on impulsive, reactive aggression do not retaliate more on the PSAP, but rather, work harder to earn money (Gan et al., 2016).

When studied in non-clinical samples, individual differences in self-control, impulsivity, aggression, and history of violence are determined by questionnaire measures or tasks in which participants may underreport these tendencies due to social desirability (Saunders, 1991). It has been proposed that in many cases, social desirability may explain the low correlations between self-reported aggression and behavioral measures of aggression (Lobbestael, 2015; Vigil-Colet et al., 2012). Therefore, to study reactive aggression in a neurotypical sample in an experimental setting, an ideal task would elicit aggression even if participants are motivated to act in a socially desirable manner.

To fill this methodological gap and allow for an empirical test of the link between impulsivity, self-control, trait-level aggression, and costly, reactive aggression, we designed a new paradigm, called the Retaliate or Carry-on: Reactive AGgression Experiment, or RC-RAGE. The RC-RAGE differs from the PSAP in that provocations are more visually salient and prolonged (thereby putting more pressure on self-control), and retaliations are very easy, immediate, and more visually violent. However, as in the PSAP (but not in the Competitive Reaction Time Task), there is a financial cost to retaliating, which creates the conflict that requires self-control. The original version of our task diverges from standard lab-based paradigms where there is an ostensible other person being harmed directly or indirectly, due to concerns over the beliefs regarding deception/cover stories (McCarthy & Elson, 2018) and due to a desire to increase the contexts in which the task can be used (e.g., outside of the lab). We propose that this version of the task (referred to as the Computer Opponent Version) can provide a proxy for impulsive, reactive aggression or a measure of reactive aggressive tendencies that allows greater use-case flexibility. However, we also developed a version of the task (referred to as the Human Opponent Version) in which participants are misled to believe that they are playing against another player, making it a more valid measure of aggression. Additionally, given concerns about the flexible measures used in quantifying aggression in paradigms such as the Competitive Reaction Time Task (Elson, 2016; Elson et al., 2014), we preregistered our experiment, measurement approach, and confirmatory analyses.

We hypothesized that participants who reported higher trait aggression, higher trait impulsivity, and poorer trait self-control would show higher levels of costly retaliation in this paradigm. Additionally, based on research linking aggression and state-level anger (Harmon-Jones & Sigelman, 2001; Denson et al., 2009), we hypothesized that angry affective states would be associated with costly retaliation. Lastly, based on our pilot data, we hypothesized that retaliation in this task would provide a measure of impulsive, costly aggression that is less affected or unaffected by participants' desires to "look good" (i.e., social desirability).

Consistent with the pre-registered hypotheses, we find that more costly retaliation is strongly linked to dispositional aggression, the tendency to act impulsively, and angry state affect, and not underestimated due to social desirability. Further, these results were found regardless of whether participants believed they were playing against a computer or another participant. Importantly, by creating a tangible, monetary incentive to ignore provocation, we show that the tendency to act impulsively and without thought is a better predictor of costly reactive aggression than other forms of self-control, such as the ability to delay gratification, resist temptation, or plan ahead. Together, these results suggest that the RC-RAGE task provides a robust measure of impulsive, costly aggression that can be used to better elucidate the factors that lead to impulsive aggression even when there is a clear incentive to ignore provocation and carry on.

Materials and methods

Experimental design

To examine whether retaliation on this novel task: 1) corresponds to individual differences in dispositional aggression, impulsivity, and self-control, and 2) provides a measure costly reactive aggression unaffected by social desirability, we had participants complete a number of questionnaires either before or after completing the RC-RAGE. The order was counterbalanced so that one-half of participants would complete the questionnaires first and the other half would complete the RC-RAGE first. To see whether costly retaliation on our task was also sensitive to current emotional state, particularly feelings of hostility, all participants completed a state affect assessment directly before performing the RC-RAGE. A working version of the original (Computer Opponent) version of this task (including all the instructions, audio checks, and practice), which can be run on any modern web browser, can be accessed here: https://kywch.github.io/RC-RAGE_jsPsych/rc-rage-demo.html. The task was programmed using jsPsych (de Leeuw, 2015), a JavaScript-based library designed for running behavioral experiments via web browser. Code for the task can be accessed at https://github.com/SCENeLabWSU/RC-RAGE.

Participants

All participants were recruited via CloudResearch (https://www.cloudresearch.com/; Litman et al., 2017) to complete the full study procedures via Amazon Mechanical Turk (MTurk). Study procedures were approved by the University of Chicago Institutional Review Board (IRB no. 14-1065). Across both versions, participants were excluded if they failed more than two attention-check questions in the questionnaires or if they completed fewer than six trials where they were provoked (as provocation could occur at one of six potential times). Additionally, due to a somewhat high rate of HITs returned due to the browser window size requirements, fewer participants completed the task than the target *N*s specified on CloudResearch.

For the confirmatory study of the Computer Opponent Version reported in this work, the targeted *N* was 378 participants and 364 participants completed all study procedures. This target *N* was specified in our pre-registration and was chosen to match the sample size from the initial version of this task where all retaliation was equally costly so that comparisons between versions could be made. Of the 364 participants collected, ten were excluded from data analysis: four were excluded due to failed attention check questions, five were excluded due to insufficient number of trials, and one participant was excluded for both of these reasons, resulting in a final *N* of 354. Per our pre-registration, we set up the HIT to have equivalent proportions of male and female participants and to restrict the age range of participants to those between 18 and 55. Of the 354 analyzed participants, 174 identified as male, 174 identified as female, four identified as non-binary or other, and two preferred not to disclose their gender. All participants were between 19 and 61 years of age[1] ($M = 36.0$, $SD = 8.6$). The order of task procedures were roughly equal, with 178 participants completing the RC-RAGE first and 176 participants completing the questionnaires first. We recruited the same number of participants and exclusion criteria specified in our pre-registration to the Human Opponent version. A total of 380 participants completed the study procedures, with 14 participants excluded from data analysis: four for missing more than two attention checks, eight for completing insufficient trials, and two for both, yielding a final sample of $N = 366$. Of the 366 participants, the range of ages was 20 to 71 ($M = 40.5$, $SD = 11.36$). 157 participants identified as female, 204 identified as male, two identified as non-binary or other, and three preferred not to disclose their gender. A total of 180 participants completed the questionnaires first and 186 completed the RC-RAGE task first.

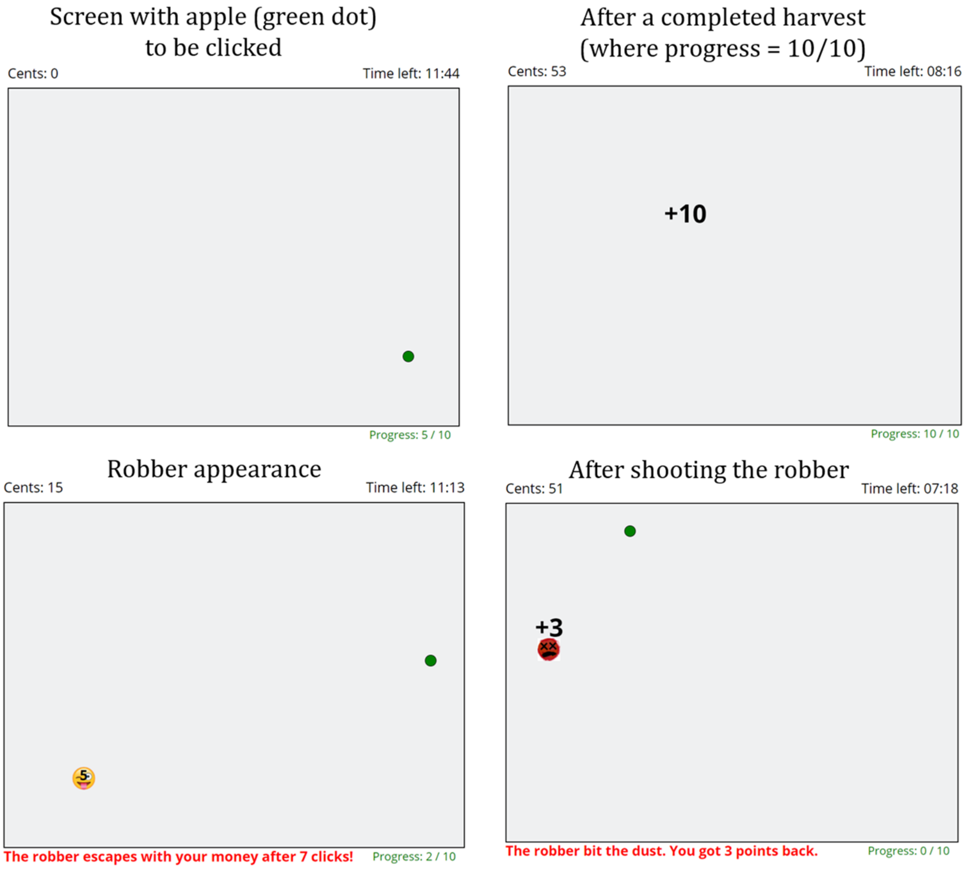
Retest sample- computer opponent version

To evaluate test–retest reliability of the RC-RAGE Computer Opponent Version, all participants who completed either the original pilot study ($N = 96$) or the confirmatory study ($N = 354$) were invited to complete the study again, either approximately 13 months later (for the confirmatory study participants) or 14 months later (for the pilot study participants). Of the potential 450 participants, 191 completed the study a second time. Due to insufficient number of completed trials, three participants were excluded from further analysis, yielding a final sample of 188 participants. Of these 188, 83 identified as female, 102 identified as male, and three identified as non-binary or other. The mean age at original testing (T1) was 38.9 years ($SD = 10.3$). At the first testing (T1), 90 participants did the questionnaires first and 98 completed the task first. At re-test (T2), 96 participants completed the questionnaires first and 92 completed the task first. As participants in the RC-RAGE Human Opponent Version were debriefed immediately regarding the deception per the requirements of our IRB, we did not conduct a test–retest analysis on this version.

Full procedure

All participants were required to pass a CAPTCHA validation before beginning. Participants whose browser windows were not sufficiently large (minimum = 1024 x 660 pixels) or did not pass the sound check (testing that experiment audio could be heard clearly) were prevented from completing further experimental procedures. Order of task and questionnaires was randomized using built-in Qualtrics functions. Regardless of order, all participants completed the PANAS directly before beginning the RC-RAGE. Between instructions, practice, and actual experiment, the RC-RAGE component took approximately 15–18 minutes. In addition to the questionnaires collected for confirmatory analysis (Questionnaires section below), participants also completed the Novaco Anger Scale (Novaco, 1994), the Selfishness Questionnaire (Raine & Uh, 2019), and the Big Five Inventory (John et al., 1999). After completing all questionnaires and the RC-RAGE, participants completed a brief demographics questionnaire. Participants in the Computer Opponent Version also answered a few questions regarding their experiences with the RC-RAGE task (i.e., how irritating they found the robber, whether they found the gunshot noise annoying or anxiety-inducing, etc., see Supplementary Materials for more details). Participants in the Human Opponent Version completed a series of open-ended suspicion probe questions from Edlund and Nichols (2019)

[1] Requested age range was specified at the level of setting up the HIT, but actual age was determined by self-reported year of birth, suggesting some participants may not be truthful about their age in their MTurk profile.



Screen with apple (green dot) to be clicked

Cents: 0

Time left: 11:44

Progress: 5 / 10

After a completed harvest (where progress = 10/10)

Cents: 53

Time left: 08:16

+10

Progress: 10 / 10

Robber appearance

Cents: 15

Time left: 11:13

The robber escapes with your money after 7 clicks!

Progress: 2 / 10

After shooting the robber

Cents: 51

Time left: 07:18

+3

The robber bit the dust. You got 3 points back.

Progress: 0 / 10

Fig. 1 Example screenshots from RC-RAGE Computer Opponent with event descriptors

to determine whether participants thought they were indeed playing against another participant (See Supplementary Materials for full list of questions). In this questionnaire, 23% of participants ($N = 85$) brought up the possibility that they were playing against a computer opponent.

The retaliate or carry-on: reactive aggression experiment (RC-RAGE)

In the RC-RAGE, participants were asked to maximize their earnings in 12 min by clicking on green dots (referred to as apples) moving around the 800 x 600-pixel game board at the center of the screen. When they clicked on an apple, it disappeared and appeared after 500 ms at a random location, which is sampled from a uniform distribution across the game board and at least 200 pixels away (if the location within the 200 pixels was sampled, sampling was repeated). Once participants clicked on ten apples in a row (i.e., a harvest), they were able to cash out and 10 cents was added to their total earnings.

Occasionally, an opponent (referred to as the "robber") would appear on the screen, steal 5 cents of their money, and remain there for some period of time. Participants could retaliate against the robber and get 3 cents back by shooting him twice to destroy him, but when they did so, they would lose their progress towards their harvest. For example, if a participant clicked on seven apples in a row, their current progress towards the harvest would be 7/10, and if the robber appeared at this point and the participant retaliated, they would get 3 cents back but their progress towards the harvest would return to 0/10. The robber always disappeared before participants could complete their progress towards the ten apples, and after he disappeared, they would lose their chance to get 3 cents back. Thus, the robber forced participants to continuously choose between whether to retaliate and lose progress or to ignore him and carry on (See Fig. 1 for Computer Opponent RC-RAGE participant interface examples; See Supplementary Materials for Human Opponent Version interface).

In the Human Opponent Version, participants were misled to believe they were randomly assigned to play the role of the “gatherer” and they would be playing against another player, who was assigned to be the robber. After reading the instructions, participants were ostensibly paired with an opponent, who started with 200 cents and who was instructed to just sit and monitor the progress of the gatherer.

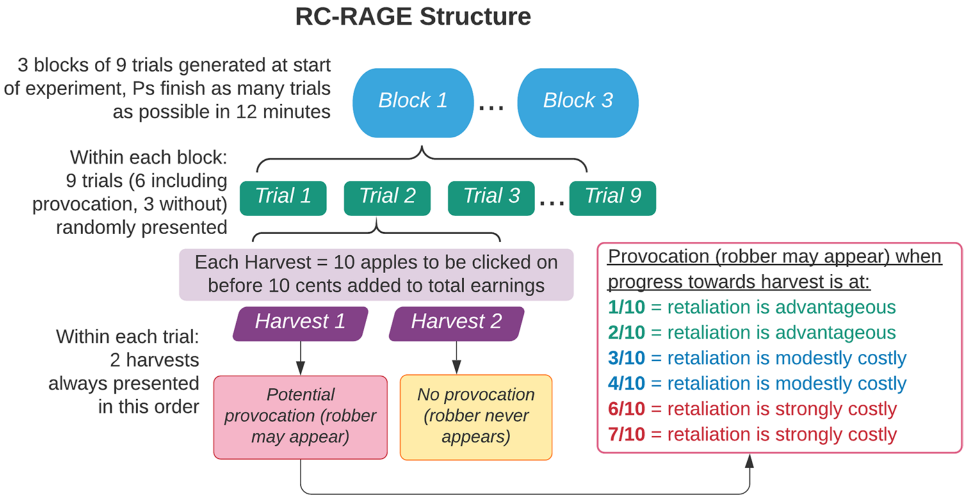

RC-RAGE Structure
3 blocks of 9 trials generated at start of experiment, Ps finish as many trials as possible in 12 minutes
Block 1
...
Block 3
Within each block: 9 trials (6 including provocation, 3 without) randomly presented
Trial 1
Trial 2
Trial 3
...
Trial 9
Each Harvest = 10 apples to be clicked on before 10 cents added to total earnings
Harvest 1
Harvest 2
Within each trial: 2 harvests always presented in this order
Potential provocation (robber may appear)
No provocation (robber never appears)
Provocation (robber may appear) when progress towards harvest is at:
1/10 = retaliation is advantageous
2/10 = retaliation is advantageous
3/10 = retaliation is modestly costly
4/10 = retaliation is modestly costly
6/10 = retaliation is strongly costly
7/10 = retaliation is strongly costly

Fig. 2 Structure of the full RC-RAGE

Participants were told the robber was occasionally given the opportunity to steal from them, and they may or may not do so when given the opportunity. In this version, both the participant's earnings ("Gatherer Bonus") and the opponent's earnings ("Robber Bonus") were displayed above the game board, and the phrase "The robber escapes with your money after X clicks" was replaced with "Your opponent will escape with your money after X clicks".

In both versions, the time at which the robber appeared was manipulated so that, depending on how much progress the participant had made towards the harvest of ten apples, the cost associated with retaliation was varied. For example, the progress lost by retaliation was greater when the robber appeared after the participant had already clicked on seven apples (progress count: 7/10) than if the participant retaliated after the participant had only clicked on one apple (progress count: 1/10). To quantify the extent to which retaliation was more or less costly, we calculated the monetary value of each mouse click during the task and compared the value of mouse clicks across conditions. To calculate the value of each mouse click, we defined a trial as consisting of two harvests (i.e., two instances of 10 apple clicks in a row) and designed the robber to always appear during the first harvest. If the robber did not appear in that trial, participants completed two harvests and earned 20 cents by clicking 20 apples without interruption, resulting in 1 cent per click (Fig. 2).

If the robber appeared when participants have only clicked one apple (i.e., progress count: 1/10) and they chose to ignore and carry on, they would earn 15 cents (two harvests - 5 cents taken by the robber) by clicking 20 apples in a row, resulting in 0.75 cents per click. If they chose to retaliate and reset their progress, they would earn 18 cents (by getting back 3 cents) by making 23 clicks (one lost click + two clicks to destroy the robber + 20 apples), resulting in 0.783 cents per click. Thus, when the robber appears at the 1/10 progress, the value of each click is slightly higher with retaliation than self-restraint (ignoring the robber and carrying on).

If the robber appeared when participants have clicked two apples (i.e., progress count: 2/10) and they choose to ignore and carry on, the value of click remained the same 0.75 cent per click (15 cents divided by 20 clicks). If they chose to retaliate and reset their progress, they would earn 18 cents by making 24 clicks (two lost clicks + two clicks to destroy the robber + 20 apples), resulting in 0.75 cents per click. Thus, when the robber appears at the 2/10 progress, the value of click is the same between retaliation and self-restraint.

If the robber appeared when participants have clicked three apples (i.e., progress count: 3/10) and they chose to retaliate and reset their progress, they would earn 18 cents by making 25 clicks (three lost clicks + two clicks to destroy the robber + 20 apples), resulting in 0.72 cents per click. If the robber appeared when the progress count was 7/10, they would earn 18 cents by making 29 clicks (seven lost clicks + two clicks to the robber + 20 apples), resulting in 0.62 cents per click. Thus, when the robber appears at 3+/10 progress, the value of click is lower with retaliation (0.72 cents per click or less) than self-restraint (0.75 cents per click).

Based on this calculation, participants were instructed that it is financially best to ignore the robber if they have already clicked on more than two apples (i.e., progress count is more than 2/10). However, if they have only clicked on one or two apples, it is financially advantageous to destroy the robber right away. In this task, the robber may appear at six different times: when progress is 1/10, 2/10, 3/10, 4/10, 6/10 or 7/10. Six possibilities were included (rather than all potential progress points, 0 to 9) to increase the likelihood

Table 1 Percent of participants retaliating at least once in each situation split by gender and order for Computer Opponent Version

	Overall (*N* = 354)	Gender			Order	
		Female (*N* = 174)	Male (*N* = 174)	Other (*N* = 6)	Survey 1st (*N* = 176)	Task 1st (*N* = 178)
Advantageous						
Did not retaliate	27 (7.6%)	12 (6.9%)	14 (8.0%)	1 (16.7%)	10 (5.7%)	17 (9.6%)
Retaliated	327 (92.4%)	162 (93.1%)	160 (92.0%)	5 (83.3%)	166 (94.3%)	161 (90.4%)
Modestly Costly						
Did not retaliate	153 (43.2%)	76 (43.7%)	73 (42.0%)	4 (66.7%)	86 (48.9%)	67 (37.6%)
Retaliated	201 (56.8%)	98 (56.3%)	101 (58.0%)	2 (33.3%)	90 (51.1%)	111 (62.4%)
Strongly Costly						
Did not retaliate	264 (74.6%)	134 (77.0%)	127 (73.0%)	3 (50.0%)	140 (79.5%)	124 (69.7%)
Retaliated	90 (25.4%)	40 (23.0%)	47 (27.0%)	3 (50.0%)	36 (20.5%)	54 (30.3%)

Table 2 Percent of participants retaliating at least once in each situation split by gender, order, and suspicion detected for Human Opponent Version

	Overall (*N* = 366)	Gender			Order		Suspicious	
		Female (*N* = 157)	Male (*N* = 204)	Other (*N* = 5)	Survey 1st (*N* = 180)	Task 1st (*N* = 186)	No (*N* = 281)	Yes (*N* = 85)
Advantageous								
Did not retaliate	14 (3.8%)	7 (4.5%)	6 (2.9%)	1 (20.0%)	9 (5.0%)	5 (2.7%)	13 (4.6%)	1 (1.2%)
Retaliated	352 (96.2%)	150 (95.5%)	198 (97.1%)	4 (80.0%)	171 (95.0%)	181 (97.3%)	268 (95.4%)	84 (98.8%)
Modestly costly								
Did not retaliate	143 (39.1%)	60 (38.2%)	80 (39.2%)	3 (60.0%)	77 (42.8%)	66 (35.5%)	109 (38.8%)	34 (40.0%)
Retaliated	223 (60.9%)	97 (61.8%)	124 (60.8%)	2 (40.0%)	103 (57.2%)	120 (64.5%)	172 (61.2%)	51 (60.0%)
Strongly costly								
Did not retaliate	246 (67.2%)	100 (63.7%)	142 (69.6%)	4 (80.0%)	119 (66.1%)	127 (68.3%)	184 (65.5%)	62 (72.9%)
Retaliated	120 (32.8%)	57 (36.3%)	62 (30.4%)	1 (20.0%)	61 (33.9%)	59 (31.7%)	97 (34.5%)	23 (27.1%)

that participants would experience each condition multiple times throughout the experiment. To ensure participants experience all conditions, these six conditions were grouped into a batch of nine trials by adding three trials in which the robber does not appear, and the order of these nine trials was randomized so that the robber appeared on two-thirds of trials. As a trial consisted of two harvests, the robber appeared on one-third of harvests (Fig. 2). Three batches of the shuffled nine trials (thus 27 trials total) were prepared for each participant to complete within 12 min. On average, participants in the Computer Opponent Version completed 20 trials (M = 19.8, SD = 3.5) with 431 clicks (M = 431.4, SD = 69.2), earning a bonus of 348 cents (M = 348.1, SD = 59.7), and participants in the Human Opponent Version completed 20 trials (M = 19.5, SD = 3.3) with 428 clicks (M = 428.3, SD = 63.3), earning a bonus of 345 cents (M = 344.7, SD = 54.8).

We operationally define retaliation at 1-or-2 clicks in as advantageous, retaliation at 3-or-4 clicks in as modestly costly, and 6-or-7 clicks in as strongly costly. These three conditions were created based on what participants were explicitly told (i.e., it was financially best to retaliate if progress is at 1 or 2 clicks in but not if they've made progress greater than 2/10 clicks) and with the goal of keeping the number of trial types per condition consistent (i.e., combining trials where progress was 3-or-4, 6-or-7). For each condition (advantageous, modestly costly, strongly costly), retaliation rate was calculated as the (#retaliations in condition / #trials in condition). The percentage of participants who retaliated at least once for each type can be found in Tables 1 and 2.

In the Computer Opponent Version, participants were taken through step-by-step instructions and given 2 min to practice before beginning the real experiment. Money earned during these 2 min did not count towards their total earnings.

In the Human Opponent Version, participants were not given an opportunity to practice as interacting with the ostensible other participant during practice could affect their behavior in the main experiment and/or could increase participant suspicion regarding the deception. For the main experiment, participants performed the RC-RAGE for 12 min, and their total earnings were credited as a cash bonus at the end of the session. During both practice and main rounds, attention checks appeared where a letter of the alphabet is auditorily presented, and participants were asked to press the alphabet key they just heard right away, as the timer continues during these attention checks. This was to ensure that participants did have their sound on and were continuously performing the task.

Questionnaires

The self-report constructs of primary interest in this work were aggression, impulsivity, self-control, and social desirability. To measure trait-level aggression, the Buss-Perry Aggression Questionnaire was used (Buss & Perry, 1992), which includes 29 statements where participants are asked to rate how characteristic each statement is of them on a scale of 1–5 (1 = extremely uncharacteristic of me, 5 = extremely characteristic of me). The BPAQ measures total aggression as well as four subscales of aggression: Physical Aggression (example statement: “If somebody hits me, I hit back”), Verbal Aggression (example statement: “I can’t help getting into arguments when people disagree with me”), Anger (example statement: “Sometimes I fly off the handle for no good reason”), and Hostility (example statement: “At times I feel like I have gotten a raw deal out of life”).

To measure impulsivity, the Barratt Impulsiveness Scale (BIS-11) was used (Patton et al., 1995). On the BIS-11, participants read 30 statements about the ways people act and think and respond on a 1–4 scale (1 = rarely/never and 4 = almost always/always) whether it applies to them. The BIS-11 generates a score for three second-order factors of impulsivity: Motor (example statement: “I do things without thinking”), Attentional (example statement: “I am restless at the theater or lectures”), and Non-planning (example statement: reverse coded “I plan tasks carefully”).

Self-control was measured using the Brief Self-Control Scale (Tangney et al., 2004), which generates a total self-control score. For each of the 13 items, participants are asked to rate whether the statement (such as “I am good at resisting temptation” or “I have a hard time breaking bad habits”) applies to them on a scale of 1-5 (1 = not at all, 5 = very much).

Socially desirable responding was measured using the Marlowe-Crowne Social Desirability Scale (MCSDS; Crowne & Marlowe, 1960). The MCSDS is a 33-item scale where participants respond with whether the statement is true or false of them. Higher total scores on this questionnaire suggests the respondent is presenting themself in an unrealistically positive manner. Sample items on the MCSDS include “I never hesitate to go out of my way to help someone in trouble” and “I’m always willing to admit it when I make a mistake.”

State affect was calculated using the short form of the Positive and Negative Affect Schedule (PANAS; Watson et al., 1988). Though overall positive and negative affect were calculated for exploratory analyses, confirmatory analyses were conducted looking specifically at ratings for the “Hostility” and “Irritability” items. Correlations between the self-report measures are shown in the Supplementary Materials.

Statistical analysis

Statistical analysis was conducted using R version 3.5.1 (R Foundation for Statistical Computing, www.rproject.org). Correlations between retaliation rates and other variables of interest were calculated using the function ‘rcorr’ in the ‘Hmisc’ package (Harrel, 2022). *P* values for confirmatory tests were Bonferroni corrected to control the family-wise error rate (alpha = 0.05/32 confirmatory correlations = 0.0015). Comparison of correlation coefficients was conducted using the function ‘cocor.dep.groups.overlap’ in the ‘cocor’ package (Diedenhofen & Musch, 2015). This specific function tests for significant differences in correlation coefficients in one group with an overlapping variable and a one-tailed alpha of 0.05 was used to test for significance.

For logistic mixed effects regressions, the function ‘glmer’ in the ‘lme4’ package (Bates et al., 2015) was used. For each self-report measure, the model was specified as:

$$\begin{aligned}&\text{glmer(Retaliation\_Rate} \sim \text{Costliness}^{*}\text{Self\_Report\_Measure}\\&\quad + (1 \mid \text{sub}), \text{family} = \text{binomial}, \text{nAGQ} = 10),\end{aligned}$$

where costliness was a categorical factor corresponding to the robber appearing at position 1-or-2 (advantageous), position 3-or-4 (modestly costly), or position 6-or-7 (strongly costly). In this model, estimates are based on an adaptive Gaussian Hermite approximation of the likelihood using ten integration points. To get the mixed effects results, a multilevel bootstrapping procedure was employed to obtain bootstrapped mean estimates and 95% confidence intervals. For each analysis, 1000 bootstrapped samples were used. Predicted probability plots were created using the ‘ggpredict’ function of the ‘ggeffects’ package (Lüdecke, 2018).

All participants in the RC-RAGE Computer Opponent Version (*N* = 354) are included in the analyses below. Results are presented separately for participants who did not suspect deception (*N* = 281) and the full sample in the RC-RAGE Human Opponent Version (*N* = 366).

Results

Confirmatory correlation results

Our first hypothesis was that participants who reported higher trait level aggression and impulsivity, and lower trait self-control would show higher levels of costly retaliation on this paradigm. However, as our retaliation rate data were highly left skewed (in the strongly costly condition) and bimodal (in the modestly costly condition), fitting a linear slope to the relationship was ultimately a suboptimal analysis. As such, we provide an overview of these results below, but much more detailed results of these confirmatory correlations can be found in the Supplementary Materials.

For trait aggression, small-to-medium positive correlations were found between all subscales of the BPAQ (Anger, Hostility, Physical Aggression, and Verbal Aggression) and both forms of costly retaliation (i.e., modestly costly and strongly costly, all r values between 0.15 and 0.31, $p < 0.006$) in the Computer Opponent Version. For trait impulsivity, medium positive correlations were found between motor impulsivity (BIS-Motor) and both forms of costly retaliation (r = 0.42 and 0.36, $p <$ 0.001). The other BIS subscales (Attentional and Nonplanning) and the Brief Self-Control Scale were significant with $p < 0.05$ uncorrected but did not survive Bonferroni correction.

For the non-suspicious participants in the Human Opponent version, small-to-medium positive correlations between strongly and modestly costly retaliation and BPAQ Anger, BPAQ Physical Aggression, and BIS Motor Impulsivity (all r values between 0.18 and 0.29) all survived Bonferroni correction ($p < 0.001$). In the full sample of the Human Opponent Version, significant positive correlations between strongly costly retaliation and BPAQ Anger and Physical Aggression, and BIS Motor Impulsivity all were significant before correction (all r values between 0.19 and 0.25, $p < 0.001$). No relationships with modestly costly retaliation survived correction. All other correlations were not significant (ps > 0.08).

Our second hypothesis was that negative affective state, particularly feelings of anger, would be associated with costly retaliation. To test this, we conducted correlations between retaliation rate and two items from the PANAS that best reflected an angry affective state: hostile and irritable. In the Computer Opponent Version, significant, positive associations were found between hostile affective and irritable affective state and both modestly costly and strongly costly retaliation (all $p \leq 0.001$). These effects were also found for hostile affect ($p < 0.001$ for strongly costly, $p \leq$ 0.007 for modestly costly) in the Human Opponent Version (both full sample and non-suspicious participants). However, they were not found for irritable affect. Taken together, these results suggest that both an angry emotional state and trait-level aggression and motor impulsivity contribute to the likelihood of engaging in costly, reactive aggression.

Effects of social desirability

Our final confirmatory hypothesis was that this task would provide a measure of impulsive aggression that would not be affected by socially desirable responding. We predicted that social desirability would negatively correlate with measures of aggression and impulsivity, positively correlate with self-control, and would be unrelated to the retaliation rate on the RC-RAGE. Detailed results can be found in the Supplementary Materials.

Across all versions of the task, significant correlations were found between social desirability and most (or all) of the BPAQ subscales, BIS-Attentional, BIS-Nonplanning, and Brief Self Control. There was not a significant correlation between social desirability and BIS-Motor Impulsivity in either version, though it was all in the expected direction. Critically, social desirability was not significantly correlated with costly retaliation in the Computer opponent version (modestly costly: r = 0.15, p = 0.006; strongly costly: r = 0.1, p = 0.068), in the Human Opponent Version non-suspicious sample (modestly costly: r = 0.03, p = 0.65; strongly costly: r = 0.15, p = 0.01), or in the Human Opponent full sample (modestly costly: r = 0.07, p = 0.20; strongly costly: r = 0.15, p = 0.005), and these correlations were positive (i.e., higher social desirability, higher rates of retaliation). These results suggest that, at least in an online context, individuals do not refrain from impulsive aggression on the RC-RAGE due to a desire to maintain socially acceptable behavior.

Comparing the relationships between impulsivity/self-control and costly aggression

The confirmatory correlation analyses suggested that there may be a stronger link between costly aggression and the tendency to act impulsively than other forms of impulsivity and self-control. To directly test whether this is the case, one sided z-tests specifically testing whether correlations with costly retaliation and BIS-Motor were larger than correlations with other measures of self-control were used. The detailed results of these correlations are presented in Table 3. In both versions of the task, motor impulsivity showed a significantly larger correlation with strongly costly retaliation rate than did attentional impulsivity, nonplanning impulsivity, and self-control (all $p < 0.001$). The same was true for modestly costly retaliation rate, where

Table 3 *Z*-tests comparing impulsivity/self-control measures and costly retaliation rate

Computer Opponent Version ($N = 354$)	z-test vs. BIS-motor ~ strongly costly RR	z-test vs. BIS-motor ~ modestly costly RR
BIS- Attentional	$z = 5.06, p < 0.001$	$z = 4.78, p < 0.001$
BIS- Nonplanning	$z = 5.06, p < 0.001$	$z = 4.61, p < 0.001$
Brief Self-Control	$z = 5.11, p < 0.001$	$z = 4.81, p < 0.001$
Human Opponent Version, non-suspicious subset ($N = 281$)	z-test vs. BIS-motor ~ strongly costly RR	z-test vs. BIS-motor ~ modestly costly RR
BIS- Attentional	$z = 3.18, p < 0.001$	$z = 3.12, p < 0.001$
BIS- Nonplanning	$z = 4.10, p < 0.001$	$z = 2.16, p = 0.015$
Brief Self-Control	$z = 5.02, p < 0.001$	$z = 3.35, p < 0.001$
Human Opponent Version, all participants ($N = 366$)	z-test vs. BIS-motor ~ strongly costly RR	z-test vs. BIS-motor ~ modestly costly RR
BIS- Attentional	$z = 3.34, p < 0.001$	$z = 1.65, p = 0.05$
BIS- Nonplanning	$z = 4.17, p < 0.001$	$z = 1.96, p = 0.025$
Brief Self-Control	$z = 4.01, p < 0.001$	$z = 2.56, p = 0.005$

larger correlations were found with motor impulsivity than other measures of impulsivity or self-control (all $p < 0.001$ in Computer Opponent Version, all $p \leq 0.01$ in Non-suspicious participants in the Human Opponent Version, all $p \leq 0.05$ in Human Opponent Version full sample). These results suggest that retaliation on the RC-RAGE where the demands on self-control are high is most tightly linked to individual differences in the tendency to act impulsively.

Predicting costliness of retaliation by self-report measure

In addition to the confirmatory correlations, exploratory logistic mixed effect regressions were conducted to examine interactions between the self-report measures and retaliation rate as a function of how costly retaliation was. This was conducted for two primary reasons. First, this approach allows for a specific examination of how the costliness of retaliation (rather than retaliation in general) relates to trait aggression, impulsivity, self-control, and angry affect. Second, as the values of retaliation rate are limited to being between 0 and 1 and do not follow a normal distribution, conducting a logistic regression is more appropriate than the linear regression used in correlations, which are presumed to follow a gaussian distribution. These models were run separately (as opposed to all-in-one regression) due to high inter-measure correlations which caused multicollinearity if included in the same model.

Detailed results for dispositional aggression can be found in Table 4 and Fig. 3. In the Computer Opponent Version, self-reported aggression as measured by all four BPAQ subscales (physical, verbal, anger, and hostility) significantly interacted with how costly retaliation was in predicting retaliation rate. More specifically, dispositional aggression showed a greater relationship with retaliation rate when it was modestly costly or strongly costly relative to when it was advantageous. This interaction effect was largest for physical aggression and anger (see Table 4 and Fig. 3). For the non-suspicious participants in the Human Opponent version, BPAQ Physical Aggression and BPAQ Anger also showed significantly stronger relationships with modestly and strong costly retaliation rates (Table 4, Fig. 3). However, the relationships were weaker with BPAQ Hostility and BPAQ Verbal Aggression in this version (interactions only significant with strongly costly retaliation vs. advantageous). A similar pattern of results was found in the full sample (See Supplementary Materials for Regression Tables).

The tendency to act impulsively (as measured by the BIS-Motor subscale) also yielded significant interactions with costliness of retaliation in predicting retaliation rate, wherein participants higher on motor impulsivity also retaliated more when it was modestly costly and strongly costly relative to advantageous (Table 4, Fig. 3). This was the case across both Human Opponent and Computer Opponent Versions of the RC-RAGE. Neither of the other BIS scales (attentional, non-planning) showed a significant interaction, nor did trait self-control as measured by the Brief Self-Control Scale (see Supplementary Figures). Across both versions of the RC-RAGE, hostile state affect showed significant interactions with costliness of retaliation rate, where higher levels of state hostility were more related to modestly costly or strongly costly retaliation than advantageous retaliation

Table 4 Logistic mixed effects regression tables with significant interactions - Computer Opponent Version & Non-Suspicious Participants in Human Opponent Version. Logistic Regression Models Predicting Retaliation Rate by Costliness Condition (Advantageous, Modestly Costly, Strongly Costly) and Self-report measures: BPAQ Physical Aggression, BPAQ Anger, BIS Motor Impulsivity, and Hostile Affect. Results for the full sample of participants in the Human Opponent version are in the Supplementary Materials. Models without significant interactions in both tasks are not shown but output is accessible on the OSF project page). Full models are reported but two-way interactions are the tests of primary interest. For costliness conditions, advantageous retaliation was used as the reference. Fixed effects results are reported as estimates (**B**) with standard errors, *z*-values, *p* values. Mixed effects values are calculated through multilevel bootstrapping to generate boot mean estimates (**B**) and 95% confidence intervals

Computer Opponent Version BPAQ Physical Aggression	Fixed Effects Est. *B* (Std. Error)	Mixed Effects Est. *B* [95% CI LL, UL]	*z*	*p*
Intercept	2.58 (0.51)	3.73 [2.21, 5.50]	5.01	< 0.001
Modestly Costly Retaliation	– 4.58 (0.64)	– 7.67 [– 9.67, – 5.66]	– 7.21	< 0.001
Strongly Costly Retaliation	– 6.60 (0.75)	– 10.2 [– 12.6, – 7.7]	– 8.86	< 0.001
BPAQ Physical Aggression	– 0.15 (0.21)	– 0.19 [– 0.90, 0.46]	– 0.74	0.46
Modestly Costly * Physical	0.72 (0.25)	1.22 [0.44, 2.00]	2.91	0.003
Strongly Costly * Physical	1.09 (0.27)	1.69 [0.80, 2.6]	4.00	< 0.001
Random Effects	Variance	Std. Dev		
Subject (*n* = 354)	1.24	1.2		
AIC	771.4			
Log Likelihood	– 378.7			
Observations	1062			
Human Opponent Version Non-Suspicious Participants BPAQ Physical Aggression	Fixed Effects Est. *B* (Std. Error)	Mixed Effects Est. *B* [95% CI LL, UL]	*z*	*p*
Intercept	3.21 (0.63)	3.76 [2.40, 5.19]	5.08	< 0.001
Modestly Costly Retaliation	– 4.94 (0.75)	– 5.82 [– 7.51, – 4.31]	– 6.56	< 0.001
Strongly Costly Retaliation	– 6.16 (0.81)	– 7.25 [– 9.12, – 5.45]	– 7.62	< 0.001
BPAQ Physical Aggression	– 0.37 (0.26)	– 0.44 [– 1.0, 0.17]	– 1.42	0.16
Modestly Costly * Physical	0.93 (0.31)	1.11 [0.49, 1.78]	3.03	0.002
Strongly Costly * Physical	1.08 (0.32)	1.27 [0.57, 2.01]	3.38	< 0.001
Random Effects	Variance	Std. Dev		
Subject (*n* = 281)	0.63	0.80		
AIC	577.0			
Log Likelihood	– 281.5			
Observations	843			
Computer Opponent Version BPAQ Anger	Fixed Effects Est. *B* (Std. Error)	Mixed Effects Est. *B* [95% CI LL, UL]	*z*	*p*
Intercept	2.13 (0.49)	3.22 [1.66, 4.85]	4.31	< 0.001
Modestly Costly Retaliation	– 4.15 (0.62)	– 6.74 [– 8.6, – 4.65]	– 6.75	< 0.001
Strongly Costly Retaliation	– 6.28 (0.73)	– 9.91 [– 12.18, – 7.74]	– 8.62	< 0.001
BPAQ Anger	0.04 (0.2)	0.05 [– 0.61, 0.73]	0.21	0.84
Modestly Costly * Anger	0.54 (0.24)	0.80 [0.00, 1.57]	2.19	0.028
Strongly Costly * Anger	0.95 (0.27)	1.56 [0.75, 2.42]	3.54	< 0.001
Random Effects	Variance	Std. Dev		
Subject (*n* = 354)	1.20	1.1		
AIC	769.2			
Log Likelihood	– 377.6			
Observations	1062			
Human Opponent Version Non-Suspicious Participants BPAQ Anger	Fixed Effects Est. *B* (Std. Error)	Mixed Effects Est. *B* [95% CI LL, UL]	*z*	*p*
Intercept	3.0 (0.58)	3.51 [2.16, 4.97]	5.15	< 0.001
Modestly Costly Retaliation	– 4.47 (0.70)	– 5.30 [– 6.80, – 3.81]	– 6.44	<0.001

Table 4 (continued)

Strongly Costly Retaliation	– 5.78 (0.75)	– 6.81 [– 8.55, – 5.14]	– 7.70	< 0.001
BPAQ Anger	– 0.28 (0.25)	– 0.33 [– 0.95, 0.31]	– 1.14	0.25
Modestly Costly * Anger	0.74 (0.29)	0.88 [0.23, 1.50]	2.52	0.011
Strongly Costly * Anger	0.94 (0.31)	1.10 [0.35, 1.84]	3.07	0.002
Random Effects	Variance	Std. Dev		
Subject (*n* = 281)	0.65	0.80		
AIC	581.9			
Log Likelihood	– 284.0			
Observations	843			
Computer Opponent Version	Fixed Effects Est.	Mixed Effects Est.	*z*	*p*
BIS Motor Impulsivity	*B* (Std. Error)	*B* [95% CI LL, UL]		
Intercept	1.99 (0.74)	2.79 [0.34, 4.93]	2.69	0.007
Modestly Costly Retaliation	– 5.81 (0.95)	– 9.43 [– 12.2, – 6.48]	– 6.14	< 0.001
Strongly Costly Retaliation	– 8.31 (1.07)	– 12.6 [– 16.0, – 9.28]	– 7.77	< 0.001
Motor Impulsivity	0.09 (0.38)	0.23 [– 0.84, 1.47]	0.23	0.81
Modestly Costly * Motor	1.53 (0.47)	2.40 [0.97, 3.78]	3.30	< 0.001
Strongly Costly * Motor	2.19 (0.50)	3.28 [1.67, 4.83]	4.40	< 0.001
Random Effects	Variance	Std. Dev		
Subject (*n* = 354)	0.98	0.99		
AIC	736.8			
Log Likelihood	– 361.4			
Observations	1062			
Human Opponent Version Non-Suspicious Participants	Fixed Effects Est.	Mixed Effects Est.	*z*	*p*
BIS Motor Impulsivity	*B* (Std. Error)	*B* [95% CI LL, UL]		
Intercept	3.75 (0.90)	4.34 [2.18, 6.44]	4.19	0.007
Modestly Costly Retaliation	– 5.84 (1.07)	– 6.80 [– 8.97, – 4.47]	– 5.43	< 0.001
Strongly Costly Retaliation	– 7.93 (1.16)	– 9.21 [– 11.6, – 6.61]	– 6.84	< 0.001
Motor Impulsivity	– 0.70 (0.44)	– 0.80 [– 1.82, 0.35]	1.59	0.11
Modestly Costly * Motor	1.52 (0.53)	1.76 [0.58, 2.80]	2.89	0.004
Strongly Costly * Motor	2.14 (0.55)	2.47 [1.23, 3.60]	3.87	< 0.001
Random Effects	Variance	Std. Dev		
Subject (*n* = 281)	0.65	0.80		
AIC	575.3			
Log Likelihood	– 280.6			
Observations	843			
Computer Opponent Version	Fixed Effects Est.	Mixed Effects Est.	*z*	*p*
Hostile Affect	*B* (Std. Error)	*B* [95% CI LL, UL]		
Intercept	2.40 (0.33)	3.38 [2.26, 4.4]4	7.16	< 0.001
Modestly Costly Retaliation	– 4.11 (0.43)	– 6.56 [– 8.01, – 5.22]	– 9.52	< 0.001
Strongly Costly Retaliation	– 5.68 (0.50)	– 8.60 [– 10.3, – 7.08]	– 11.37	< 0.001
Hostile Affect	– 0.13 (0.19)	– 0.04 [– 0.63, 0.67]	– 0.69	0.49
Modestly Costly * Hostile	0.86 (0.24)	1.19 [0.38, 1.95]	3.64	< 0.001
Strongly Costly * Hostile	1.13 (0.24)	1.66 [0.81, 2.50]	4.72	< 0.001
Random Effects	Variance	Std. Dev		
Subject (*n* = 354)	1.12	1.09		
AIC	751.5			
Log Likelihood	– 368.7			
Observations	1062			

Table 4 (continued)

Human Opponent Version Non-Suspicious Participants Hostile Affect	Fixed Effects Est. *B* (Std. Error)	Mixed Effects Est. *B* [95% CI LL, UL]	*z*	*p*
Intercept	3.01 (0.39)	3.48 [2.62, 4.36]	7.64	< 0.001
Modestly Costly Retaliation	– 4.10 (0.48)	– 4.79 [– 5.76, – 3.80]	– 8.56	< 0.001
Strongly Costly Retaliation	– 5.32 (0.52)	– 6.23 [– 7.37, – 5.10]	– 10.14	< 0.001
Hostile Affect	– 0.40 (0.20)	– 0.45 [– 0.91, 0.12]	– 2.02	0.044
Modestly Costly * Hostile	0.81 (0.25)	0.95 [0.41, 1.49]	3.23	0.001
Strongly Costly * Hostile	1.03 (0.25)	1.2 [0.58, 1.76]	4.07	<0.001
Random Effects	Variance	Std. Dev		
Subject (*n* = 281)	0.71	0.84		
AIC	577.4			
Log Likelihood	– 281.7			
Observations	843			

(Fig. 3, Table 4). There was also a significant interaction with irritable affect in the Computer Opponent Version but not the Human Opponent Version (See Supplementary Materials).

Test–retest reliability analysis for computer opponent version

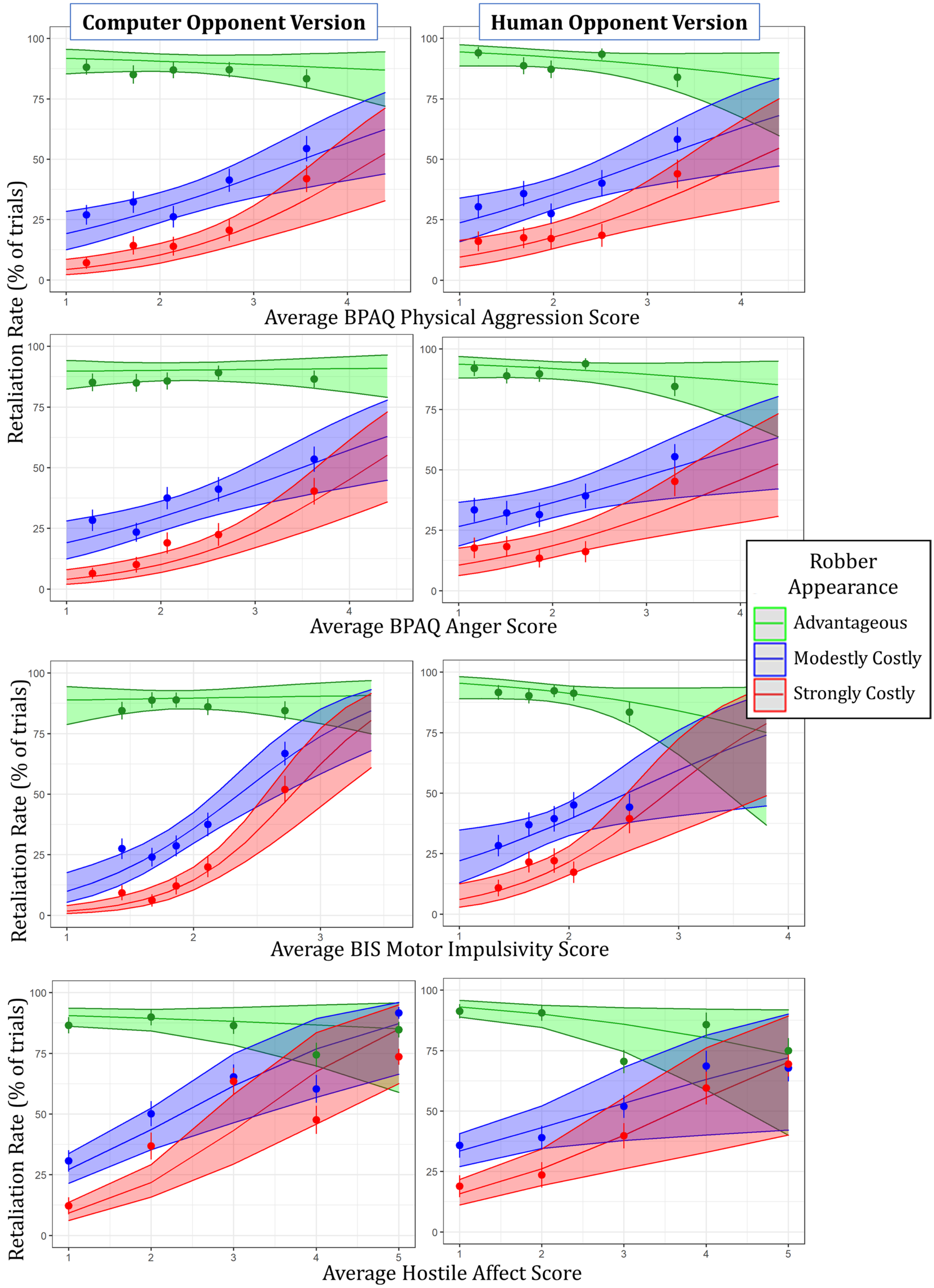
To evaluate the reliability of the Computer Opponent Version of the task, correlations between costly retaliation rates at original testing (T1) and at retesting (T2) were first conducted for the 188 participants who completed both sessions. For modestly costly retaliation rates, the test–retest correlation was $r = 0.55$, and for strongly costly, the correlation was 0.54. These correlations are quite high given that we do expect state-level factors to influence retaliation rates, and as there was a long time between testing (13 to 14 months).

Correlations between retaliation rates and self-report measures were also compared in this sample at T1 and T2. Overall, this sub-sample showed lower levels of costly retaliation, hostile affect, and trait aggression relative to the full sample but similar levels of impulsivity and self-control. (See Test–Retest Supplement for more details on these differences). Too few participants in this sample engaged in strongly costly retaliation at T2 (n = 24 out of 188) for statistical analyses to be meaningful, so the correlations reported focus on modestly costly retaliation rates.

In this sample, BIS-Motor impulsivity showed small-to-medium correlations with modestly costly retaliation at both time points ($r = 0.32$ at T1, $r = 0.21$ at T2). For aggression, the overall correlations were lower than in the full sample as there were overall fewer participants scoring high on trait aggression, but all were in the same direction (higher aggression and higher modestly costly retaliation rates). For BPAQ Anger, the correlation with modestly costly retaliation was $r = 0.14$ at T1 and $r = 0.11$ at T2. For BPAQ Physical Aggression, the correlation at T1 was $r = 0.12$ and $r = 0.1$ at T2. For BPAQ Verbal Aggression, the correlation at T1 was $r = 0.05$ and $r = 0.18$ at T2. BPAQ Hostility showed relatively low correlations at both time points: $r = 0.06$ at T1 and $r = 0.08$ at T2. State affect also showed weak correlations at T1 and T2 in this sample (all r between 0.01 and 0.05), again likely due to overall very low rates of irritable or hostile state affect in the sub-sample, particularly at T2.

Discussion

The link between self-control, impulsivity, and aggression has been hypothesized for decades, influencing key theories of criminal behavior and psychiatric disorders (Best et al., 2002; Gottfredson & Hirschi, 1990; Siever, 2008). However, a more nuanced understanding of how impulsivity and impaired self-control might lead to aggression upon provocation requires a metric of reactive aggression that: 1) can be used in an experimental setting, 2) allows for immediate, impulsive responding, and 3) has a tangible cost associated with retaliation, thereby creating a conflict between desired and financially optimal responding, and 4) is not influenced by socially desirable responding. The RC-RAGE was designed to fill this methodological gap, and the results of our pre-registered, confirmatory analyses showed that costly retaliation was linked to trait-level aggression, the tendency to act impulsively, and angry state affect, but was not negatively related to social desirability. Subsequent regressions demonstrated that in each of these cases, the relationships were stronger as the financial disadvantage to retaliating increased. This pattern of effects was replicated in a version of the RC-RAGE which led participants to believe they were playing against another human, where retaliating against the robber accurately reflects the definition of aggression. Further, the effect sizes were larger when examined in participants who did believe they were playing against another


Computer Opponent Version
Human Opponent Version
Retaliation Rate (% of trials)
100
75
50
25
0
1
2
3
4
5
Average BPAQ Physical Aggression Score
Average BPAQ Anger Score
Average BIS Motor Impulsivity Score
Average Hostile Affect Score
Robber Appearance
Advantageous
Modestly Costly
Strongly Costly

◀**Fig. 3** Predicted probabilities + averaged data for retaliation rate as a function of costliness (retaliation type) and dispositional measures. Results are presented separately for the Computer Opponent Version (left) and the non-suspicious participants in the Human Opponent Version (right). *Line graphs* represent predicted responses of fixed effects from logistic regression models presented in Table 4. *Shaded areas* indicate standard errors on the predicted probabilities. *Point-range plots* reflect actual data and were calculated by splitting participants into five groups based on self-report measures. In the point-range plots, *dots* represent the mean retaliation by robber appearance within groups and *bars* represent standard errors of the means. For BPAQ and BIS, the five groups were calculated using quintiles. For hostile affect, the raw values (1–5) were used. Across both versions, BPAQ Physical Aggression and Anger subscales generated significant interactions for both modestly costly and strongly costly retaliation relative to advantageous retaliation. For average BPAQ measures, the range of responses is 1–5, with higher values indicating greater aggression. Only BIS Motor Impulsivity generated significant interactions for both modestly and strongly costly retaliation relative to advantageous retaliation. For averaged BIS scores, the range of responses is 1–4 with higher values indicating greater impulsivity. Hostile Affect showed significant interactions with modestly and strongly costly retaliation. For Hostile Affect, the range of responses is 1–5 with higher values indicating greater feelings of hostility

person vs. the full sample of participants (including those who suspected their opponent was not real). In addition to the experimental utility provided by this task, the results of this work provide support for the idea that the tendency to act impulsively and without thinking can lead to reactive aggression upon provocation despite clearly stated incentives to inhibit an aggressive response.

As hypothesized, costly retaliation on the RC-RAGE was positively related to dispositional aggression. The results of these relationships were either equivalent to or larger in magnitude to the correlations typically found using the point subtraction aggression paradigm in non-clinical populations (Geniole et al., 2017; McCloskey et al., 2009), which demonstrates that the external validity of this task is on par with other laboratory aggression measures.

This work showed a robust link between the tendency to act impulsively (i.e., motor impulsivity) and costly aggression. However, trait self-control and other forms of impulsivity (attentional impulsivity and non-planning) were not significantly related in the multivariable logistic regressions, suggesting that costly, reactive aggression is strongly influenced by the tendency to act impulsively, but less related to the tendency to plan ahead or delay gratification. While the ability to delay gratification, resist temptation, or persevere on tasks are important elements of self-control more broadly, the results of *z*-tests comparing these correlations suggest that reactive aggression upon provocation is more specifically linked to the tendency to act on impulse. This finding is consistent with the neurobiological frameworks of reactive aggression, which suggests that aggression will occur when there is a mismatch between the urge to retaliate driven by subcortical, limbic regions and inhibition of this desired action by prefrontal cortical regions (Davidson et al., 2000; Siever, 2008).

Consistent with research demonstrating the link between angry emotional states and aggression (Beames et al., 2020; Denson et al., 2009), we found that costly retaliation was also related to hostile affect. The more hostile the affective state, the more likely a participant was to retaliate when it was financially costly but not when it was financially advantageous. While this result does suggest that affective state may also influence costly reactive aggression, it is also possible that those participants reporting high state anger are also more hostile on a dispositional/trait level. Previous work has demonstrated a strong link between trait anger and reactive aggression (Wilkowski & Robinson, 2010), and as trait anger and hostility (as measured by the BPAQ) were correlated with state hostility in our sample, the extent to which this effect is primarily driven by current emotional state cannot be readily determined. In subsequent studies, it would be interesting to experimentally induce feelings of anger to better elucidate the role that state affect plays in costly reactive aggression.

Furthermore, we found interactions between dispositional aggression, the tendency to act impulsively, state anger and the costliness of retaliation. This feature is important, as it suggests that it's not simply the retaliation response itself that is associated with trait aggression, impulsivity, or anger on our task, but rather, the relationship is with financially costly retaliation that ought to be inhibited. That is, in the situation where a person is provoked, those high on trait aggression, motor impulsivity, and state anger are more likely to disregard the explicit financial incentives to ignore the provocation, choosing instead to retaliate despite a cost, which is suboptimal from a financial perspective.

It's worth noting that, while aggression is generally perceived as socially undesirable behavior, there are still scenarios in which choosing to behave aggressively may be incentivized, as in the case of instrumental aggression (where inflicting harm may be more of a means to an end) or in the case of a competition (i.e., behaving aggressively in a sport or competition; Taylor, 1967). It is generally agreed upon that there may be multiple motives to act aggressively in any given situation and many researchers now argue against a simple dichotomization of instrumental vs. hostile aggressive motives (Allen & Anderson, 2017; Bushman & Anderson, 2001). However, to understand the role of self-control and impulsivity in reactive aggression, the current task was designed to specifically create situations where participants were instructed to ignore provocation and use self-control processes to carry on with the task at hand. This work demonstrates that the RC-RAGE task is effective in identifying trait and state predictors of impulsive aggression upon provocation where the goals of the task (earn money) and the desirable response (retaliation) are at odds.

Our study contains a few notable limitations. Unlike most other laboratory-based aggression tasks where participants are ostensibly told they are harming another real individual, in the Computer Opponent Version of the RC-RAGE, participants are not led to believe that the opponent ("robber") is indeed another human that they are harming. This version may be preferable if experimenters would prefer to measure aggressive tendencies and avoid cover stories and clever experimental set-ups (i.e., the person experiencing harm is in "another room"), particularly if this is being conducted in a non-traditional context. This approach also mitigates the question of whether participants either "buy" the deception and whether participants are simply responding in accordance with what the experimenter wants (McCarthy & Elson, 2018). Approximately 23% of our participants indicated they had doubts about whether the other player in the Human Opponent version was, indeed, another human participant. This may be an unacceptable level of attrition for some researchers, who may prefer to measure a proxy of aggression or a measure of aggressive tendencies without the potential loss of data due to suspicious participants.

With the Computer Opponent Version, we aimed to create a task that could be used in a greater variety of settings, such as an online environment or alternative testing site, which would allow the experiment to be conducted in much larger and more diverse samples than previous aggression studies. This would also afford flexibility for potential environmental manipulations to look at some of the physical environmental effects on reactive aggression (i.e., in a place with natural scenery vs. urban scenery; Kuo & Sullivan, 2001). However, if the primary aim of the researchers is to gather a measure of impulsive, reactive aggression that ostensibly harms another person (and, therefore, has greater external validity), the Human Opponent Version would likely be the better one to use. Ultimately, as costly retaliation on both versions relates to dispositional aggression, the tendency to act impulsively, and angry affect, researchers may be able to choose which version best fits their particular use case.

A potential limitation of this task relative to other similar aggression tasks is that, to manipulate the costliness of retaliation, our experimental design (and the strategy recommended to participants) required them to learn the rule about when it was financially best to retaliate and when it was not. While this recommendation was stated explicitly, it is possible that some participants retaliated more or less often because they did not read the instructions to learn the recommended strategy. This is unlikely to be the driving force in our results as we removed participants who demonstrated inattentiveness in other parts of the study. However, future iterations of this study may wish to include an additional manipulation check where they asked participants to re-state the recommended strategy.

It should also be noted that we likely had some non-random attrition in our study sample for the test–retest analysis. Specifically, participants who were less likely to retaliate in T1 were more likely to return to complete the task at T2. This creates some limitations in the interpretability of our test–retest analysis, and future work may try to use a shorter window or additional incentive structures to reduce attrition rates. Lastly, while our confirmatory analyses supported the hypothesis that participants would not avoid retaliating on the RC-RAGE due to concerns over self-presentation (social desirability), it remains possible that this may not translate to an in-person, less anonymous context. Future work would be needed to determine how context-dependent this effect is.

Additionally, although this study identified both state and trait predictors of costly, reactive aggression, it did not examine the effects of longer-term situational or environmental factors. This is important as physical and social environmental influences have been identified as key determinants of self-control and decision-making processes (Sheehy-Skeffington, 2020). This emerging body of research demonstrates that it is not only the short-term situational context that influences self-control, impulsivity, and aggression, but chronic exposure to environmental stressors, structural prejudice against groups, financial and environmental instability, and the effects of low socio-economic status can have major impacts on aggressive behavior (Figueredo et al., 2020; Lawson et al., 2018; Sheehy-Skeffington, 2020). Therefore, a key next step in this research would be to examine how longer-term environmental effects may relate to impulsivity and costly, reactive aggression on the RC-RAGE task.

In summary, the current work introduces two versions of a novel experimental paradigm to test the trait and state-level predictors of costly impulsive, reactive aggression. By including an immediate retaliation option and making aggression costly, the RC-RAGE places high demands on self-control and allows for impulsive responses. The results demonstrated that the tendency towards acting impulsively was more predictive of costly retaliation than other types of self-control, and suggest that while self-control (broadly defined) can predict a variety of aggressive or antisocial behaviors (Gottfredson & Hirschi, 1990), costly reactive aggression is best predicted by impulsivity. This effect is also consistent with neurobiological theories of aggression (Coccaro et al., 2011; Nelson & Trainor, 2007; Siever, 2008), and an exciting future direction for this work would be to use this paradigm to disentangle the contributions of reactivity to provocation (driven by limbic regions) and impaired impulse inhibition (subserved by prefrontal cortical regions) in an experimental setting. Going forward, this work also provides a more general opportunity for future research on further elucidating the state-level, trait-level, and environmental predictors of costly reactive aggression upon provocation.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13428-023-02066-9.

Acknowledgments We thank Nak Won Rim for his assistance in generating Figure S2 and all the members of the Environmental Neuroscience Lab who tested out the task during the development stage.

Code availability • Rmarkdown analysis code and output can also be found on the OSF page for this project: https://osf.io/796rs/

Author contribution
Conceptualization: Kimberly L. Meidenbauer and Kyoung W. Choe.
Data Curation: Kimberly L. Meidenbauer and Kyoung W. Choe.
Formal Analysis: Kimberly L. Meidenbauer, Kyoung W. Choe, Akram Bakkour, Michael Inzlicht, and Marc G. Berman.
Funding Acquisition: Kimberly L. Meidenbauer & Marc G. Berman.
Investigation: Kimberly L. Meidenbauer and Kyoung W. Choe.
Methodology: Kimberly L. Meidenbauer and Kyoung W. Choe.
Project Administration: Marc G. Berman.
Resources: Kimberly L. Meidenbauer, Kyoung W. Choe, Michael L. Meidenbauer
Software: Kimberly L. Meidenbauer, Kyoung W. Choe, Michael L. Meidenbauer.
Supervision: Marc G. Berman.
Validation: Kimberly L. Meidenbauer.
Visualization: Kimberly L. Meidenbauer.
Writing - Original Draft Preparation: Kimberly L. Meidenbauer.
Writing - Review & Editing: Kimberly L. Meidenbauer, Kyoung W. Choe, Akram Bakkour, Michael Inzlicht, and Marc G. Berman.

Funding The current work was supported in part by National Science Foundation Smart and Connected Cities grant S&CC1952050 (awarded to MGB) and in part by the University of Chicago Center for Health Administration Studies Seed Grant (awarded to MGB and KLM).

Data availability

- Data from confirmatory studies (present work) and pilot study are available at the project OSF page: https://osf.io/796rs/
- Pre-registration for this study is available at: https://osf.io/czn43/
- A working version of the Computer Opponent Version of the task is accessible at https://kywch.github.io/RC-RAGE_jsPsych/rc-rage-demo.html
- The jsPsych code for the RC-RAGE can be accessed at https://github.com/SCENeLabWSU/RC-RAGE

Declarations

Conflicts of interest The authors declare no conflicts of interest.

Ethics Approval All procedures were approved by the University of Chicago Institutional Review Board (IRB No. 14-1065).

Consent to participate Informed consent was obtained from all individual participants included in the study.

Consent for publication N/A.

References

Allen, J. J., & Anderson, C. A. (2017). General Aggression Model. In *The International Encyclopedia of Media Effects* (pp. 1–15). Wiley. https://doi.org/10.1002/9781118783764.wbieme0078

Barratt, E. S., Stanford, M. S., Dowdy, L., Liebman, M. J., & Kent, T. A. (1999). Impulsive and premeditated aggression: A factor analysis of self-reported acts. *Psychiatry Research, 86*(2), 163–173. https://doi.org/10.1016/s0165-1781(99)00024-4

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, Articles, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beames, J. R., Gilam, G., Schofield, T. P., Schira, M. M., & Denson, T. F. (2020). The impact of self-control training on neural responses following anger provocation. *Social Neuroscience*, 1–13. https://doi.org/10.1080/17470919.2020.1799860

Berkman, E. T., Hutcherson, C. A., Livingston, J. L., Kahn, L. E., & Inzlicht, M. (2017). Self-Control as Value-Based Choice. *Current Directions in Psychological Science, 26*(5), 422–428. https://doi.org/10.1177/0963721417704394

Best, M., Williams, J. M., & Coccaro, E. F. (2002). Evidence for a dysfunctional prefrontal circuit in patients with an impulsive aggressive disorder. *Proceedings of the National Academy of Sciences of the United States of America, 99*(12), 8448–8453. https://doi.org/10.1073/pnas.112604099

Boureau, Y.-L., Sokol-Hessner, P., & Daw, N. D. (2015). Deciding How To Decide: Self-Control and Meta-Decision Making. *Trends in Cognitive Sciences, 19*(11), 700–710. https://doi.org/10.1016/j.tics.2015.08.013

Bushman, B. J., & Anderson, C. A. (2001). Is it time to pull the plug on the hostile versus instrumental aggression dichotomy? *Psychological Review, 108*(1), 273–279. https://doi.org/10.1037/0033-295x.108.1.273

Buss, A. H. (1961). *The Psychology of Aggression*. John Wiley & Sons, Inc. https://doi.org/10.1037/11160-000

Buss, A. H., & Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology, 63*(3), 452–459. https://doi.org/10.1037//0022-3514.63.3.452

Cherek, D. R., Schnapp, W., Moeller, F. G., & Dougherty, D. M. (1996). Laboratory measures of aggressive responding in male parolees with violent and nonviolent histories. *Aggressive Behavior, 22*(1), 27–36. https://doi.org/10.1002/(SICI)1098-2337(1996)22:1<27::AID-AB3>3.0.CO;2-R

Cherek, D. R., Lane, S. D., Dougherty, D. M., Moeller, F. G., & White, S. (2000). Laboratory and questionnaire measures of aggression among female parolees with violent or nonviolent histories. *Aggressive Behavior, 26*(4), 291–307. https://doi.org/10.1002/1098-2337(2000)26:4<291::AID-AB2>3.0.CO;2-9

Coccaro, E. F., Sripada, C. S., Yanowitch, R. N., & Phan, K. L. (2011). Corticolimbic function in impulsive aggressive behavior. *Biological Psychiatry, 69*(12), 1153–1159. https://doi.org/10.1016/j.biopsych.2011.02.032

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349–354. https://doi.org/10.1037/h0047358

da Cunha-Bang, S., Fisher, P. M., Hjordt, L. V., Perfalk, E., Persson Skibsted, A., Bock, C., Ohlhues Baandrup, A., Deen, M., Thomsen, C., Sestoft, D. M., & Knudsen, G. M. (2017). Violent offenders respond to provocations with high amygdala and striatal reactivity. *Social Cognitive and Affective Neuroscience, 12*(5), 802–810. https://doi.org/10.1093/scan/nsx006

Davidson, R. J., Putnam, K. M., & Larson, C. L. (2000). Dysfunction in the neural circuitry of emotion regulation--a possible prelude to violence. *Science, 289*(5479), 591–594. https://doi.org/10.1126/science.289.5479.591

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods, 47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Denson, T. F., Pedersen, W. C., Ronquillo, J., & Nandy, A. S. (2009). The angry brain: Neural correlates of anger, angry rumination, and aggressive personality. *Journal of Cognitive Neuroscience, 21*(4), 734–744. https://doi.org/10.1162/jocn.2009.21051

Denson, T. F., DeWall, C. N., & Finkel, E. J. (2012). Self-Control and Aggression. *Current Directions in Psychological Science, 21*(1), 20–25. https://doi.org/10.1177/0963721411429451

DeWall, C. N., Anderson, C. A., & Bushman, B. J. (2011). The general aggression model: Theoretical extensions to violence. *Psychology of Violence, 1*(3), 245–258. https://doi.org/10.1037/a0023842

Diedenhofen, B., & Musch, J. (2015). cocor: A comprehensive solution for the statistical comparison of correlations. *PloS One, 10*(3), e0121945. https://doi.org/10.1371/journal.pone.0121945

Edlund, J. E., & Nichols, A. L. (2019). *Advanced Research Methods for the Social and Behavioral Sciences*. Cambridge University Press. https://doi.org/10.1017/9781108349383

Eisenberg, I. W., Bissett, P. G., Zeynep Enkavi, A., Li, J., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Uncovering the structure of self-regulation through data-driven ontology discovery. *Nature Communications, 10*(1), 2319. https://doi.org/10.1038/s41467-019-10301-1

Elson, M. (2016). FlexibleMeasures.com: Competitive reaction time task. https://doi.org/10.17605/OSF.IO/4G7FV

Elson, M., Mohseni, M. R., Breuer, J., Scharkow, M., & Quandt, T. (2014). Press CRTT to measure aggressive behavior: The unstandardized use of the competitive reaction time task in aggression research. *Psychological Assessment, 26*(2), 419–432. https://doi.org/10.1037/a0035569

Figueredo, A. J., Black, C. J., Patch, E. A., Heym, N., Ferreira, J. H. B. P., Varella, M. A. C., Defelipe, R. P., Cosentino, L. A. M., Castro, F. N., Natividade, J. C., Hattori, W. T., Pérez-Ramos, M., Madison, G., & Fernandes, H. B. F. (2020). The cascade of chaos: From early adversity to interpersonal aggression. *Evolutionary Behavioral Sciences*. https://doi.org/10.1037/ebs0000241

Finkel, E. J., DeWall, C. N., Slotter, E. B., McNulty, J. K., Pond, R. S., Jr., & Atkins, D. C. (2012). Using I^3 theory to clarify when dispositional aggressiveness predicts intimate partner violence perpetration. *Journal of Personality and Social Psychology, 102*(3), 533–549. https://doi.org/10.1037/a0025651

Gan, G., Preston-Campbell, R. N., Moeller, S. J., Steinberg, J. L., Lane, S. D., Maloney, T., Parvaz, M. A., Goldstein, R. Z., & Alia-Klein, N. (2016). Reward vs. Retaliation-the Role of the Mesocorticolimbic Salience Network in Human Reactive Aggression. *Frontiers in Behavioral Neuroscience, 10*, 179. https://doi.org/10.3389/fnbeh.2016.00179

García-Forero, C., Gallardo-Pujol, D., Maydeu-Olivares, A., & Andrés-Pueyo, A. (2009). Disentangling impulsiveness, aggressiveness and impulsive aggression: An empirical approach using self-report measures. *Psychiatry Research, 168*(1), 40–49. https://doi.org/10.1016/j.psychres.2008.04.002

Geniole, S. N., MacDonell, E. T., & McCormick, C. M. (2017). The Point Subtraction Aggression Paradigm as a laboratory tool for investigating the neuroendocrinology of aggression and competition. *Hormones and Behavior, 92*, 103–116. https://doi.org/10.1016/j.yhbeh.2016.04.006

Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime*. Stanford University Press. http://www.sup.org/books/title/?id=2686

Harrell, F. E., Jr (2022). Hmisc: Harrell miscellaneous. https://CRAN.R-project.org/package=Hmisc

Harmon-Jones, E., & Sigelman, J. (2001). State anger and prefrontal brain activity: Evidence that insult-related relative left-prefrontal activation is associated with experienced anger and aggression. *Journal of Personality and Social Psychology, 80*(5), 797–803. https://doi.org/10.1037/0022-3514.80.5.797

Hofmann, W., Friese, M., & Strack, F. (2009). Impulse and Self-Control From a Dual-Systems Perspective. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science, 4*(2), 162–176. https://doi.org/10.1111/j.1745-6924.2009.01116.x

Inzlicht, M., & Schmeichel, B. J. (2012). What Is Ego Depletion? Toward a Mechanistic Revision of the Resource Model of Self-Control. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science, 7*(5), 450–463. https://doi.org/10.1177/1745691612454134

Inzlicht, M., Werner, K. M., Briskin, J. L., & Roberts, B. W. (2021). Integrating Models of Self-Regulation. *Annual Review of Psychology, 72*, 319–345. https://doi.org/10.1146/annurev-psych-061020-105721

John, O. P., Srivastava, S., & Others. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of Personality: Theory and Research, 2*(1999), 102–138.

Kuo, F. E., & Sullivan, W. C. (2001). Aggression and Violence in the Inner City: Effects of Environment via Mental Fatigue. *Environment and Behavior, 33*(4), 543–571. https://doi.org/10.1177/00139160121973124

Lawson, G. M., Hook, C. J., & Farah, M. J. (2018). A meta-analysis of the relationship between socioeconomic status and executive function performance among children. *Developmental Science, 21*(2), e12529. https://doi.org/10.1111/desc.12529

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods, 49*(2), 433–442. https://doi.org/10.3758/s13428-016-0727-z

Lobbestael, J. (2015). Challenges in aggression assessment: The gap between self-report and behavior, and a call for new valid behavioral paradigms. *Journal of Socialomics, 5*(01), 5–6. https://doi.org/10.4172/2167-0358.1000141

Lovallo, W. R. (2013). Early life adversity reduces stress reactivity and enhances impulsive behavior: Implications for health behaviors. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology, 90*(1), 8–16. https://doi.org/10.1016/j.ijpsycho.2012.10.006

Lüdecke, D. (2018). Ggeffects: Tidy data frames of marginal effects from regression models. *Journal of Open Source Software, 3*(26), 772. https://doi.org/10.21105/joss.00772

McCarthy, R. J., & Elson, M. (2018). A conceptual review of lab-based aggression paradigms. *Collabra. Psychology, 4*(1), 4. https://doi.org/10.1525/collabra.104

McCloskey, M. S., New, A. S., Siever, L. J., Goodman, M., Koenigsberg, H. W., Flory, J. D., & Coccaro, E. F. (2009). Evaluation of behavioral impulsivity and aggression tasks as endophenotypes for borderline personality disorder. *Journal of Psychiatric Research, 43*(12), 1036–1048. https://doi.org/10.1016/j.jpsychires.2009.01.002

Nelson, R. J., & Trainor, B. C. (2007). Neural mechanisms of aggression. *Nature Reviews. Neuroscience, 8*(7), 536–546. https://doi.org/10.1038/nrn2174

Novaco, R. W. (1994). Anger as a risk factor for violence among the mentally disordered. *Violence and Mental Disorder: Developments in Risk Assessment, 21*, 59.

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt impulsiveness scale. *Journal of Clinical Psychology, 51*(6), 768–774. https://doi.org/10.1002/1097-4679(199511)51:6<768::aid-jclp2270510607>3.0.co;2-1

Raine, A. (2008). From genes to brain to antisocial behavior. *Current Directions in Psychological Science, 17*(5), 323–328. https://doi.org/10.1111/j.1467-8721.2008.00599.x

Raine, A., & Uh, S. (2019). The Selfishness Questionnaire: Egocentric, Adaptive, and Pathological Forms of Selfishness. *Journal of Personality Assessment, 101*(5), 503–514. https://doi.org/10.1080/00223891.2018.1455692

Ritter, D., & Eslea, M. (2005). Hot Sauce, toy guns, and graffiti: A critical account of current laboratory aggression paradigms. *Aggressive Behavior, 31*(5), 407–419. https://doi.org/10.1002/ab.20066

Saunders, D. G. (1991). Procedures for Adjusting Self-Reports of Violence for Social Desirability Bias. *Journal of Interpersonal Violence, 6*(3), 336–344. https://doi.org/10.1177/088626091006003006

Sheehy-Skeffington, J. (2020). The effects of low socioeconomic status on decision-making processes. *Current Opinion in Psychology, 33*, 183–188. https://doi.org/10.1016/j.copsyc.2019.07.043

Siever, L. J. (2008). Neurobiology of aggression and violence. *The American Journal of Psychiatry, 165*(4), 429–442. https://doi.org/10.1176/appi.ajp.2008.07111774

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality, 72*(2), 271–324. https://doi.org/10.1111/j.0022-3506.2004.00263.x

Taylor, S. P. (1967). Aggressive behavior and physiological arousal as a function of provocation and the tendency to inhibit aggression. *Journal of Personality, 35*(2), 297–310. https://doi.org/10.1111/j.1467-6494.1967.tb01430.x

Tedeschi, J. T., & Quigley, B. M. (1996). Limitations of laboratory paradigms for studying aggression. *Aggression and Violent Behavior, 1*(2), 163–177. https://doi.org/10.1016/1359-1789(95)00014-3

Vazsonyi, A. T., Mikuška, J., & Kelley, E. L. (2017). It's time: A meta-analysis on the self-control-deviance link. *Journal of Criminal Justice, 48*, 48–63. https://doi.org/10.1016/j.jcrimjus.2016.10.001

Vigil-Colet, A., Ruiz-Pamies, M., Anguiano-Carrasco, C., & Lorenzo-Seva, U. (2012). The impact of social desirability on psychometric measures of aggression. *Psicothema, 24*(2), 310–315 http://www.psicothema.com/pdf/4016.pdf

Walters, G. D. (2008). Self-Report Measures of Psychopathy, Antisocial Personality, and Criminal Lifestyle: Testing and Validating a Two-Dimensional Model. *Criminal Justice and Behavior, 35*(12), 1459–1483. https://doi.org/10.1177/0093854808320922

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Wilkowski, B. M., & Robinson, M. D. (2010). The anatomy of anger: an integrative cognitive model of trait anger and reactive aggression. *Journal of Personality, 78*(1), 9–38. https://doi.org/10.1111/j.1467-6494.2009.00607.x

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.