Facing Enemies.

Modulation of Revenge Interactions based on Opponent State Indicators of Suffering.

Inaugural-Dissertation

zur Erlangung der Doktorwürde der Fakultät für Humanwissenschaften

der Julius-Maximilians-Universität Würzburg

Vorgelegt von

Vanessa Mitschke,

Würzburg, 2022

Erstgutachter: Professor Dr. Andreas B. Eder

Zweitgutachter: Professor Dr. Wilfried Kunde

Tag des Kolloquiums: 02.12.2022

Table of contents

Summary

Research on revenge often treats vengeful acts as singular one-way experiences, an approach which fails to account for the social nature and functions of revenge. This dissertation aims to integrate emotional punishment reactions into dynamic revenge sequences to investigate the affective and cognitive consequences of revenge within a social interaction.

Exacting revenge can evoke intense affective consequences, from feelings of guilt to the genuine enjoyment of the suffering of others. In Chapter 2, affective responses towards suffering opponents and the regulation of aggression based on the appraisal of distinct suffering indicators were investigated. Results indicate that the observation of opponent pain evokes positive affect (measured via facial muscle contractions during the observation), which is followed by a downregulation of subsequent punishment. Both, positive affective reactions and the downregulation of punishment, were only observed following pain and not sadness expressions. Empathic distress, indexed by negative affective reactions, was only present following the observation of pain in non-provoking opponents. Showcasing the modulation of empathy related processes due to provocation and competition.

In Chapter 3, a significant escalation of punishment, when being confronted with Schadenfreude, was observed. Results are interpreted as supporting the assumption that opponent

monitoring processes inform subsequent action selection. The observation of opponent smiles led to imitation behavior (facial mimicry), which was partially attenuated due to previous provocation. The different functions of smile mimicry in the context of the aggressive competitive setting are discussed as containing simulation aspects (to aid in opponent understanding) and as a potential mirroring of dominance gestures, to avoid submission.

In an additional series of studies, which are presented in Chapter 4, changes in memory of opponent faces following vengeful encounters were measured. Based on provocation, and punishment outcomes (pain & anger), face memory was distorted, resulting in more positive representations of opponents that expressed pain. These results are discussed as evidence of the impact of outcome appraisals in the formation of opponent representations and are theorized to aid empathy avoidance in future interactions.

The comparison of desired and observed opponent states, is theorized to result in appraisals of the punishment outcomes, which evoke affective states, inform the action selection of subsequent punishments, and are integrated into the representation of the opponent in memory.

Overall, the results indicate that suffering cues that are congruent with the chosen punishment action are appraised as positive, evoking an increase in positive affect. The emergence of positive affect during the observation of successful aggressive actions supports recent theories about the chronification of aggressive behavior based on reinforcement learning. To allow positive affect to emerge, affective empathic responses, such as distress, are theorized to be suppressed to facilitate the goal attainment process. The suffering of the opponent constitutes the proximate goal during revenge taking, which highlights the importance of a theoretical differentiation of proximate and ultimate goals in revenge to allow for a deeper understanding of the underlying motives of complex revenge behavior.

Zusammenfassung

Die bisherige Forschung zu den Mechanismen von Racheverhalten behandelt die Ausübung von Rache häufig als eine singuläre und unilaterale Erfahrung. Diese Herangehensweise berücksichtigt den sozialen Kontext und die Funktion von Rache nur unzureichend. Die vorliegende Dissertation zielt darauf ab, mittels empirischer Befunde emotionale Reaktionen in dynamische Rachesequenzen zu integrieren.

Die Ausübung von Rache kann intensive affektive Zustände wie Schuldgefühle bis hin zu dem Empfinden von Freude über das Leid Anderer (Schadenfreude) auslösen. In Kapitel 2 werden Ergebnisse geschildert, welche die unterschiedlichen affektiven Reaktionen des Aggressors in Abhängigkeit zu Indikatoren des Leides des Gegners untersuchen. Im Rahmen der durchgeführten Studien wurde eine positive Reaktion (gemessen anhand von fazialen Muskelaktivierungen) während der Beobachtung von Schmerzausdrücken des Gegners beobachtet. Die positive affektive Reaktion geht mit einer Reduktion der darauffolgenden Strafe einher. Sowohl der positive Affekt als auch die Reduktion von Strafe wurden nur in Folge von Schmerzindikatoren und nicht nach Ausdrücken von Trauer beobachtet. Empathischer Stress nach ausgeführten Rachehandlungen, indiziert durch negative affektive Reaktionen, wurde nur gezeigt, nachdem nicht-provokative Gegner Schmerzreaktionen zeigten. Dies zeigt die Modulation von

empathischen Prozessen in Abhängigkeit von Provokation und Konkurrenz.

Kapitel 3 beschreibt die signifikante Eskalation von aggressiven Interaktionen durch eine Konfrontation mit Schadenfreude des Gegners während der Bestrafung. Dies lässt darauf schließen, dass die Beobachtung des Gegnerzustandes während der gesamten Interaktion die Wahl von Handlungen beeinflusst. Die Beobachtung des Lächelns beim Gegenüber führt zu Nachahmungsverhalten (faziale Mimikry), welches durch vorherige Provokation graduell abgeschwächt wird. Die verschiedenen Funktionen der Mimikry des Lächelns im Kontext einer aggressiv kompetitiven Situation enthalten Elemente von Simulation, zum besseren Verständnis der Gegnerreaktion, und dienen dazu als potenzielle Reaktion auf Dominanzverhalten Submission zu vermeiden.

In einer weiteren Experimentalreihe, beschrieben in Kapitel 4, wird die Integration von erlebter Provokation und daraus resultierenden empfundenen Ärger in die Erinnerung an den Gegner demonstriert. Ergebnisse dieser Studienreine legen nahe, dass die Repräsentation des Gesichtes des Gegners in der Erinnerung basierend auf vorheriger Provokation und distinkten emotionalen gegnerischen Reaktionen auf Strafe (Schmerz & Ärger) systematisch variiert. Gegner, welche Schmerzreaktionen zeigten, wurden mit positiveren Gesichtskonfigurationen repräsentiert, im Gegensatz zu Gegnern, welche Ärger zeigten. Diese Ergebnisse werden als Belege für den Einfluss von

Ergebnisbewertungen auf die Formation von Gegnerrepräsentationen und ihren potenziellen Nutzen für die Vermeidung von Empathie in zukünftigen Auseinandersetzungen diskutiert.

Aus dem Abgleich zwischen dem gewünschten und beobachteten gegnerischen Reaktionen resultiert eine Bewertung des Erfolgs der Strafe. Dieser Bewertungsvorgang kann positive Affekte auslösen, die als Information in die Wahl der nächsten Handlung einfließen und die Repräsentation des Gegners in der Erinnerung verändern.

Die Ergebnisse der hier dargestellten Studien deuten darauf hin, dass Signale des Leidens, die unmittelbar kongruent zu den zuvor gewählten Strafhandlungen sind, als positiv bewertet werden und somit zu einem unmittelbar erlebten positiven Affekt führen. Das Vorhandensein von positivem Affekt während der Beobachtung des leidenden Gegners nach erfolgreicher Bestrafung stimmt mit neueren Forschungsansätzen überein, die Verstärkungslernen als zentralen Prozess bei der Chronifizierung von aggressivem Verhalten postulieren. Das Auftreten von positivem Affekt wird durch die Herabsetzung von affektiv empathischen Prozessen gegenüber dem Gegner ermöglicht. Die Unterdrückung von affektiver Empathie erleichtert hierbei den Prozess der Zielerreichung. Das Leid des Gegners wird als proximales Ziel der Rachehandlung diskutiert. Eine Unterscheidung zwischen proximalen und distalen Zielen bei der

Ausübung von Rache ermöglicht ein tieferes Verständnis für die zugrundeliegende Motive komplexen Racheverhaltens.

CHAPTER I

Chapter 1 - On the complex nature of revenge

Revenge is one of the most prominent themes in culture, and literature and its psychological underpinnings have intrigued many scholars. Revenge behaviors can range from simple tit-for-tat punishments to complex lifelong blood feuds, resulting in often devastating consequences for everyone involved. To understand the underpinnings of revenge escalation and the processes that balance the subjective perception, punishment choice and affective consequences, vengeful behaviors have been studied on many levels of abstraction. From diary studies, assessing interpersonal consequences of revenge (Yoshimura, 2007), the subjective and often biased perception of harm (Baumeister et al., 1990), to neuroimaging studies showcasing the vast processes involved in the experience of revenge interactions (see Nelson & Trainor, 2007, for a review). To establish a starting point for the following studies, I will first provide a definition and consecutive confinement of revenge as an aggressive action.

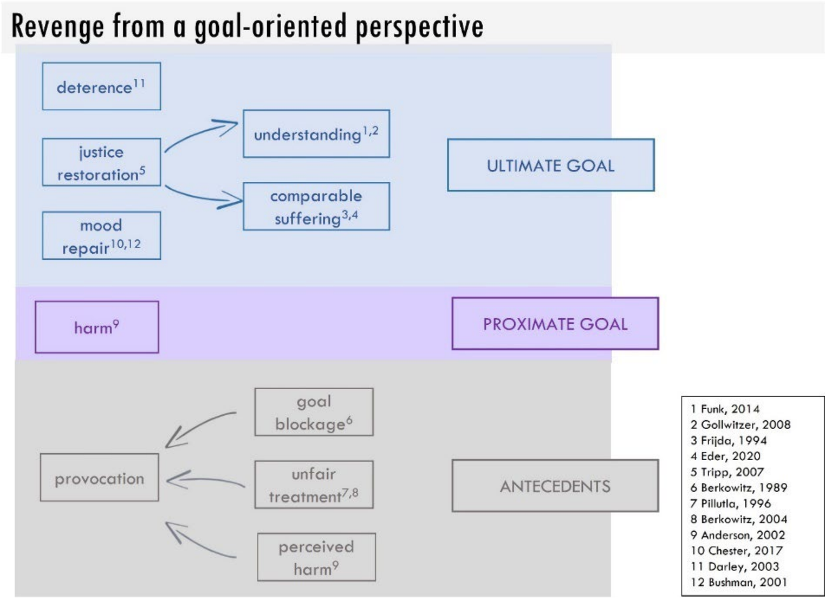
Many revenge acts can be integrated into the main framework of aggressive behavior, mainly reactive aggression. An action is defined as aggressive when it is intentionally carried out with the proximate goal of causing harm to another person, who is motivated to avoid that harm (DeWall et al., 2013). Reactively aggressive actions can be defined as aggressive actions that were preceded by acts that are subjectively evaluated as harmful and intentional (Anderson & Bushman, 2002). What the harm that is inflicted entails ranges from the loss of monetary resources, physical violence, to verbal insults affecting self-esteem or external image. In the realm of this dissertation, I will discuss harm as a negative affective state, which in our case is expressed via negative facial expressions, serving as indicators of negative affect. Viewing revenge within the framework of reactive aggression as a goal-directed behavior highlights the importance of the evaluation of action outcomes via monitoring the opponent for potential harm cues in order to adjust the consecutive aggressive actions. During the beginning of this introduction, I will review the existing literature, summarizing the potential motivations behind revenge and the complex affective states present before, during and following acts of revenge, as a basis for the experimental studies described in chapters two to four.

1.1 Motives and the Evaluation of Revenge Outcomes

Many scholars have tried to determine the motives and principles guiding revenge behavior. As already stated during the definition of aggressive acts, harm is defined as an intentional goal of aggressive actions. While harm may serve as a proximate goal during the revenge episode, the ultimate goal of revenge may differ from the desire to inflict harm. This assumption is an important addition to the original dichotomy between reactive (also called hostile) and instrumental aggression. Reactive aggression used to be defined as an impulsive (Berkowitz, 1974) and by some scholars, interestingly, goal-free behavior (Blair, 2016), separated from goal-driven – hence instrumental (derived from the Latin word "instrumentum" meaning tool) aggression. The theoretical separation between these two aggression systems has long been debated (Bushman & Anderson, 2001), since it fails to account for more complex forms of aggressive behavior, such as incidents of planned or delayed revenge for example. To accomplish even greater punishment and consequently revenge, participants high in aggressive tendencies have been shown to delay reactive aggressive behavior (West et al., 2022). This intentional delay of aggressive responding further supports the notion that revenge is not solely driven by impulsive quick decisions, and certainly not always free of the establishment of more ultimate revenge goals.

Complex revenge behaviors may involve reactive aggressive actions (provoked by unfair treatment, thus eliciting anger), in line with the proximate goal to harm the original offender, but are not limited to this proximate goal stage. Complex revenge behavior may include several higher order ultimate goals, such as deterrence (Darley & Pittman, 2003), restoration of justice (Osgood, 2017; Tripp et al., 2007) or mood repair (Bushman et al., 2001; Roberton et al., 2012). For a graphical overview please also see Figure 1. Many of these ultimate goals could potentially be obtained via harming the original offender.

Figure 1



Justice restoration, especially, as an ultimate goal of revenge, has obtained a large amount of empirical support (Carlsmith et al., 2008; Gollwitzer & Denzler, 2009; Miller, 2001; Osgood, 2017; Tripp et al., 2007). Justice sensitivity has been implicated as a stronger predictor for self-reported aggression on all sub facets (verbal, physical, relational) than hostile attribution bias (Bondü, 2018), with individuals that are highly sensitive to injustice reporting higher levels of aggressive behavior. People with a strong belief in a just world, for example, have been shown to endorse revenge more strongly than people who show a lower focus on justice (Kaiser et al., 2004). Restoring justice may be an ultimate desired end state instigating revenge, but it is less understood how this state is reached via aggressive actions and how these are appraised.

The outcomes of punishment behavior (hence the changes in opponent affect) should be monitored and evaluated with respect to the proximate goal of harm, which in turn may potentially serve higher-order goals such as justice restoration. The focal point of this dissertation lies on the consequences of the observation of opponent states indicating potential harm during the revenge episode. Due to the assumption that harm is a proximate goal underlying most ultimate goals of revenge, suffering cues will be utilized as desired consequences of the punishment choices made during the revenge episode.

The regulation of subsequent aggressive actions via the monitoring of harm cues was shown by Eder and colleagues (2020) during a set of studies employing emotional opponent displays following punishment. The authors provided avengers with opponent video reactions indicating the affective reaction towards being punished by the opponent. These videos consisted of pre-recorded emotional facial reactions (anger, sadness, pain, calm as control). The authors report changes in punishment via deducting the chosen punishment preceding the emotional displays from the punishment chosen directly after observing the opponent reactions. Results across multiple experiments indicated a downregulation of punishment following the observation of opponent pain, but no significant change in punishment following displays of anger, sadness or calm reactions. The downregulation of punishment was only slightly modulated by written indications of pain via visual analog scales, highlighting the importance of the visual observation of the pain expression. This line of research indicates that the observation of opponent states modulates punishment decision making in the subsequent interaction trial. The authors propose that viewing the opponent in pain is line with the personal goal of revenge, causing harm, hence observing the expected opponent states leads to a reduction in punishment.

The importance of the affective evaluation of harm cues is strongly influenced by the "*suffering hypothesis*", which was first introduced by Frijda (1994) in his influential essay on the desire for revenge. The author discusses the so called "Lex Talionis" an ancient punishment rule dating back to Babylonian times. It entails the idea that punishment will be evaluated based on the affective consequences for the target, which should be proportionate to the original offence that preceded the desire for punishment in the first place, based on the affective consequences suffered by the original victim. The idea of proportional suffering as a guideline for punishment can be dated back to biblical verses like "an eye for eye and a tooth for a tooth"" (English Standard Version Bible, Lev. 24:19–21). In the case of suffering as an ultimate goal of reciprocal punishment, observing pain in the opponent could be associated with balancing injustice. What remains to be largely speculative is how the balance of justice is restored. Frijda proposes that the subjective imbalance is mainly driven by affect and the desire for revenge is diminished when the original offender suffers in a comparable way.

Apart from comparable suffering, which emphasizes the importance of the affective determinants of revenge desires, comparable punishment could also be evaluated based on tit-for-tat rules (Krämer et al., 2007), as a simple rule of fairness. However, research on the escalation (gradual increase in aggressive actions) of vengeful interactions highlights the occurrence of a negative reciprocity series, each increasing the previous punishment (Shergill et al., 2003). These escalation cycles are partly due to biases in the individual evaluation of the severity of the punishment consequences (DeWall et al., 2011). In a qualitative study that sampled the everyday experience of vengeful acts (Stillwell et al., 2008), participants often attributed the goal of revenge against them as causing them pain or making them feel the same way as the interaction partner. These lay attributions seem to be grounded in the occurrence of suffering, but lay people also attribute revenge to causing an emotional trigger that makes way for behavioral change.

The idea of comparable suffering as the main desired outcome of revenge desires is not uncontested. Although part of the original definition of aggressive acts, harming the opponent in an emotional or physical manner, suffering alone does not always lead to more satisfaction with the outcome of the interaction (Funk et al., 2014). Increases in justice satisfaction have been observed following the indication of understanding by the original offender. The so called "*understanding hypothesis*" states that revenge interactions that result in offenders signaling an understanding of their wrongdoing and the subsequent punishment are subjectively experienced as more satisfying (Funk et al., 2014; Gollwitzer & Denzler, 2009).

The desire for revenge is in this case fulfilled by "*sending a message*" towards the offender (Gollwitzer & Denzler, 2009) and not by suffering alone. The authors employed a monetary provocation design where provocation is carried out via financial unfair behavior. The interaction partner would keep 90% of the shared pool of money, an unfair distribution, disadvantaging the participant. This money distribution task was then followed by a lexical decision task (LDT) including aggression related words, as a measure of revenge desire. The authors than either informed participants that their interaction partner had to do an additional task that was described as unpleasant (involving disgust related images; fate condition) or gave participants the chance to punish the interaction partner with the task (revenge condition). Thirty-four percent of the sample chose to take revenge via assigning the interaction partner to the additional task. This was followed by a chance to write down a message to the interaction partner, to which she/he either responded with understanding the reason for the punishment (unfair money distribution) or just complaining about the additional task (no understanding). Following these interactions participants completed a second round of the LDT. Results indicated that LDT scores only decreased if revenge was taken, and the offender signaled understanding of the punishment. The use of an LDT measure differs largely from the broad aggression literature, as it allows for investigation into the accessibility of aggression related concepts but does not constitute a measurement of actual aggressive behavior.

Although Gollwitzer and colleagues designed their studies to test suffering and understanding as ultimate goals directly against each other, one can still combine both approaches in a meaningful way. Goal satisfaction as observed in Gollwitzer et al. (2009) only took place after taking revenge (choosing to punish the other, followed by an indication of suffering) and an indication of understanding the reason for suffering. Nevertheless, a role of suffering cannot be ruled out, since the authors did not include a condition without a suffering outcome of the additional task. Suffering therefore may constitute a prerequisite of the revenge goal fulfillment. Punishment of the transgressor and subsequent suffering, has also been indicated as a prerequisite of forgiveness (Strelan, 2018), with insincere apologies even fostering further aggressive behavior (Miller, 2001).

Apart from the measurement, the operationalization of suffering employed by Denzler and Gollwitzer (2009) differs on multiple levels from the approach used by Eder and colleagues (2020). Both revenge goals (suffering and understanding) build on the presence of opponent reactions and monitoring thereof. Denzler and colleagues (2009) chose verbal messages to communicate suffering and understanding, while Eder at al. employed emotional facial displays in the form of opponent videos. Additionally, Eder and colleagues paired visual analogue scales as suffering indicators with facial displays of anger and pain. The scale-based indication of suffering did not add to the reduction of punishment if paired with anger and did not result in an additive effect when paired with pain. The knowledge of opponent suffering did not affect the subsequent punishment decision or goal fulfillment, but the observation of pain facial displays did, indicating potential differences between the perception of verbal / written pain cues and the visual observation of facial displays of pain.

The strong signaling qualities of facial pain displays have been studied with respect to empathy for pain. Observation of others in pain correlates with an embodied simulation of pain (Goubert et al., 2005; Singer et al., 2004; Sun et al., 2015). The observation of pain in others has been shown to involve mainly the affective parts of pain processing and does not involve the whole pain matrix (which consists of sensory and affective processing of pain cues). Hence, observing pain cues in social interactions, potentially leads to an estimation of the affective consequences for others, but not a full representation of the aversive stimulus and its consequences for one's own body. This process is believed to aid the prediction of emotional states in the other person, as well as predictions of subsequent reactions, for which a detailed processing of the aversive event in itself is not necessary.

The observation of pain cues in unfair opponents as compared to fair ones, can lead to a relative decrease in empathic processing of the target (for male observers only) during the observation of punishment (Singer et al., 2006). Interestingly, the authors report that the decrease in empathy related processing was accompanied by a relative increase in reward-based processing during the outcome phase. A shift towards reward-based processing while decreasing emotional opponent processing may facilitate aggressive actions in lowering the empathic concern for the opponent, suggesting that an avoidance of empathy may be a crucial process during the monitoring of the revenge outcomes. Hence, in the realm of this dissertation, empathy related responding, in the form of distress (negative affective reaction towards opponent suffering) and the propensity of facial mimicry of opponent reactions were investigated to allow for a deeper integration of empathic processes into aggressive behavior.

Empathy related constructs, such as mentalizing, were indicated as a crucial processing stage during the punishment decision making phase during revenge taking (Krämer et al., 2007; Lotze et al., 2007). The authors speculate that during the choice of punishment participants take not only their own aversive state but also consequences for the opponent state into account. Comparing contrasts between provoked and unprovoked conditions, increased activations of the mediofrontal gyrus were found during the decision phase following provocation. When contrasting high and low punishment choices following high provocation, Krämer and colleagues (2007), observed a significantly higher activation in the dorsal striatum before high punishment choices, an area that has been indicated in the processing of social rewards (de Quervain, 2004). Interestingly, the authors did not observe any significantly higher activation of the dorsal striatum during the observation of the punishment outcome and therefore hypothesize that not the actual outcome itself is processed as rewarding, but that reward is anticipated during the punishment selection phase. In contrast to

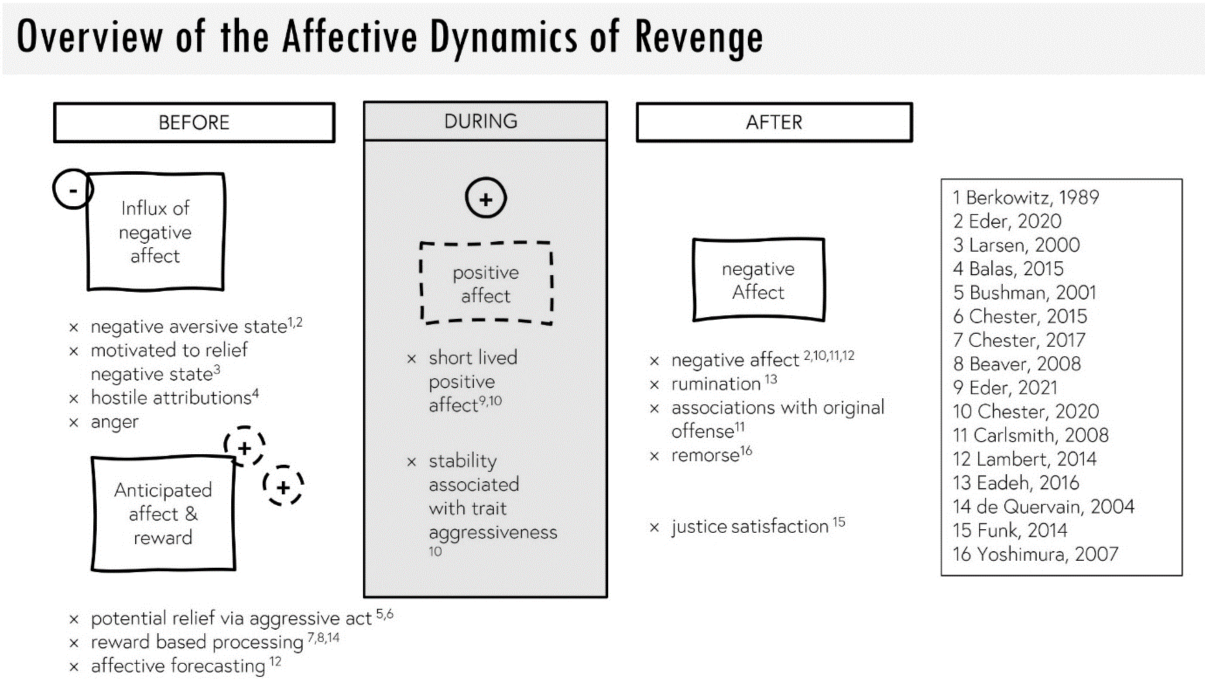
Eder et al. (2020), outcomes in the aforementioned study were only indicated via written win and loss indications and a visual cue (lightning bold) that the opponent was punished. If the direct observation of opponent states plays an important role in the assessment of punishment success, the investigated neural processes may only allow for an incomplete investigation of reward-based processing following punishment. Therefore, the addition of the measurement of affective states, potentially signaling positive affect following rewarding outcomes, or negative distress responses following undesired outcomes, provides an important approach to disentangle the mechanisms behind punishment decisions.

1.2 The Affective Dynamics of Revenge

The affective consequences of revenge have been shown to range from the anticipation of positive affect (Bushman et al., 2001; Chester, 2017), due to for example beliefs in venting anger and catharsis, to an increase in negative affect following the revenge act (Carlsmith et al., 2008; Eder et al., 2020). The complexity of affect following revenge episodes was discussed by Eadeh et al. (2017) and formalized as their *Bittersweet model* of revenge, which states that revenge acts evoke both, negative (bitter) and positive (sweet) affective states. In the following paragraph I will summarize the potential processes eliciting positive and negative affect during the revenge episode, according to their presumed timing from establishing a revenge desire to after revenge was taken. For a graphical summary of studies and potential processes that are discussed, please also see Figure 2.

Traditional aggression research mainly focused on the negative affective states that drive the desire to carry out aggressive acts, such as anger or frustration. Provocation, unfair treatment (Berkowitz & Harmon-Jones, 2004; Pillutla & Murnighan, 1996), or blocking of goals as in frustration (Berkowitz, 1989) increase feelings of anger, resulting in an aversive negative state for the mistreated individual. This negative state can be viewed as the starting point of the revenge episode. Feelings of anger should increase the desire to act out on aggressive impulses and hence further aggressive responding (Wilkowski & Robinson, 2010). Highly aversive emotional states also evoke the desire to relieve the experienced negativity (R. J. Larsen, 2000). The revenge episode starts with a subjective experience of wrongdoing and harm, resulting in an increase in anger, which needs to be resolved and dealt with. At this point a desire for revenge may be established.

Figure 2


Overview of the Affective Dynamics of Revenge
BEFORE
DURING
AFTER
-
Influx of negative affect
× negative aversive state[1,2]
× motivated to relief negative state[3]
× hostile attributions[4]
× anger
+
+
Anticipated affect & reward
× potential relief via aggressive act [5,6]
× reward based processing [7,8,14]
× affective forecasting [12]
+
positive affect
× short lived positive affect[9,10]
× stability associated with trait aggressiveness [10]
negative Affect
× negative affect [2,10,11,12]
× rumination [13]
× associations with original offense[11]
× remorse[16]
× justice satisfaction [15]
1 Berkowitz, 1989
2 Eder, 2020
3 Larsen, 2000
4 Balas, 2015
5 Bushman, 2001
6 Chester, 2015
7 Chester, 2017
8 Beaver, 2008
9 Eder, 2021
10 Chester, 2020
11 Carlsmith, 2008
12 Lambert, 2014
13 Eadeh, 2016
14 de Quervain, 2004
15 Funk, 2014
16 Yoshimura, 2007

While in a negative state, the individual seeks to relieve that state. To regulate their current affective state, the individual may choose an active form of emotion regulation. One often-discussed strategy to relieve anger is grounded in *catharsis* beliefs. In the case of anger, laypeople often hold the belief that venting (aka releasing) anger leads to a reduction in anger.

Evidence regarding catharsis beliefs indicates that while it is often anticipated that acting out on anger impulses results in relief, the contrary is actually the case (Bushman, 2002; Bushman et al., 2001). Although mood repair efforts via venting anger may not lead to the desired relief, a comparison of studies reporting catharsis beliefs and subsequent aggressive behavior indicate that if catharsis beliefs are present (Bushman et al., 2001), and anger focus is high (Gollwitzer & Bushman, 2012), individuals more readily vent their anger as a possible emotion regulation strategy.

To summarize, before the revenge episode enters the phase of choosing a concrete aggressive action, the overall state of the individual is supposedly negative with varying levels of anticipation of relief following the aggressive act. Neuroscientific studies conducted during the decision-making phase indicate a neuronal shift towards brain areas associated with potential reward estimation (Bertsch et al., 2020; Chester, 2017; de Quervain, 2004; Fanning et al., 2017).

Chester and colleagues (2017) investigated the rewarding properties of revenge during a competitive reaction time task (CRTT) focusing on activations of the Nuccleus Accumbens

(NAcc), an area related to reward-based processing. The authors measured activations during the decision phase, while participants chose the punishment volume and intensity. They observed larger NAcc activation during decisions that were preceded by high provocation by their opponent. The magnitude of the increase in reward-based processing was also correlated to self-reported aggressive tendencies in everyday life. These tendencies were measured via the Angry Mood Improvement Inventory (AMII; Bushman et al., 2001), and focused on the tendency to voice anger in an outwardly fashion. The authors reason that decision-making during the punishment phase is driven by the desire to decrease the negative affect, experienced during the provocation, via focusing on potentially positive and rewarding outcomes associated with the enactment of revenge.

Although potential reward and positive affect are expected, a wide range of evidence indicates that acts of revenge are followed by the experience of overall negative affect (Carlsmith et al., 2008; Chester et al., 2021; Eadeh et al., 2017; Eder et al., 2020; Lambert et al., 2014), contrary to the popular belief that restoring fairness or enacting punishment are inherently positive endeavors (Bushman et al., 2001). Recent empirical studies provide a more nuanced view, straying away from a strict duality of valence. Acts of revenge have been shown to be accompanied by short bursts of positive affect (Chester et al., 2021; Eder et al., 2021), potentially relieving some of the negative affect experienced due to the original offense. Nevertheless, the overall affective experience following revenge stays mostly negative.

Eadeh and colleagues (2017) investigated the affective dynamics spanning from before the enactment of revenge to after the act is carried out. Participants tend to overestimate the positive affective outcomes of revenge taking when asked to provide an estimation in an affective forecasting manner (Lambert et al., 2014). The failure to accurately predict the impact of a future event on one's own emotions has been shown with regards to the longevity of feelings (Ayton et al., 2007), as well as the overall intensity, with most participants overestimating the positive impact of an event in their personal emotional experience (Gilbert et al., 1998).

Eadah and colleagues (2017) suggested that acts of revenge are followed by periods of rumination about the original offense, hence more negative affect is experienced due to increased engagement with the original offence. Rumination has been correlated to emotionally aversive states such as sadness and anxiety (for a review see Kirkegaard Thomsen, 2006). An alternative explanation of the increase of negative affect following revenge acts could also be the experience of remorse or guilt (Yoshimura, 2007). Qualitative surveys about revenge within close relationships indicate that after anger, remorse is the most often named emotion in response to the recall of personal revenge stories. Additionally, responsibility for the suffering of another person, even if it is supposed to be beneficial or for a greater good, such as improved learning ability, has been shown to result in increases in negative affect and simulation of pain displays via facial mimicry (Lepron et al., 2014).

Although potentially reminding oneself of the original negative offense, the punishment of deserving opponents (after provocation) has been shown to be more pleasant than the punishment of undeserving opponents (Eder et al., 2021). Nevertheless, the author stress that the results indicated that overall, punishing any opponent was less pleasant than not punishing anyone. The aforementioned study did not utilize opponent state feedback, but focused on the successful choice and administration of punishment.

Positive affective reactions during aggressive acts have been shown to be associated with trait aggressiveness and more intense aggressive behavior (Chester et al., 2021). Chester and colleagues presented evidence from 8 different studies, conducted with laboratory and online aggression paradigms. Focusing on the fluctuation of affect rather than just measuring overall experience, it was shown that especially the within-person variability of positive affect across the aggressive act (i.e., before, during and after) correlates negatively with aggressive tendencies. Whereas low trait-aggressive individuals showed a significant decrease in positive affect during the aggressive act, high trait-aggressive individuals showed stable positive affect across the aggressive act. The authors split the reactive aggressive interaction into measurement phases before, during and after the aggressive act. While focusing on the intraindividual and inter individual differences in valence, the experienced intensity, and rapid shifts between valence and or intensity. This approach was chosen to focus on the instability/ stability in affective states, adding a temporal component to the investigation of affect during revenge. However, the authors exclusively employed self-reports of affect at multiple time points, an approach that is susceptible to multiple measurement difficulties.

The indication of affect via self-report scales has important methodological disadvantages. First, affective states need to be intense enough so that the participant is aware of them and hence able but also willing to report. Especially in the case of reporting schadenfreude or joy during or after the aggressive event, participants might be aware of the moral and social implications of their report and resort to underreporting. Social desirability especially affects reports of aggressive acts that were carried out by oneself (Vigil-Colet et al., 2012). Additionally, the explicit evaluation and attribution of the currently felt emotion can potentially increase the likelihood of corrective processes, resulting in a stronger focus on norms and beliefs (Gawronski & Bodenhausen, 2006). An explicit evaluation, via self-reported affect, should therefore be avoided during the observation of punishment outcomes, if possible.

To separate the affective experience after the revenge act, which could possibly be tainted by retrospective bias (Lambert et al., 2014) and or be driven by triggered negative affect based on associations with the original offense (Carlsmith et al., 2008), from the experience during the observation of revenge outcomes (opponent punishment reactions), research lines 1 and 2, measured affective responding during the observation of the punishment outcome in an unobtrusive and timely accurate manner via facial muscle electromyography (fEMG). The measurement of short-lived affective states during the interaction requires approaches that are fast and unobtrusive, to avoid interference with the revenge interaction. These research lines are mainly concerned with the immediate affective responses during the observation. Research line 3 provides are more detailed focus on longer term consequences of provocation and punishment outcomes on person memory and visual representations of opponent faces.

The overarching aim of this dissertation is to conceptualize revenge as an interactive cycle that involves the monitoring of punishment outcomes (opponent reactions towards punishment) to inform the avenger of the effect of their chosen punishment and consequently adapt their revenge behavior (punishment in consecutive round).

In the course of the next three chapters I will present sets of studies designed to investigate spontaneous facial muscle responses towards opponent emotional displays with opponents varying in prior provocation, as well as one study line investigating the integration of emotional opponent displays into visual mental opponent representations. These studies set out to shed light on the modulation of aggressive behavior via punishment outcome displays, which indicate the emotional state of the opponent, in measuring subsequent behavioral changes but also immediate affective reactions towards the outcome via fEMG, as well as the propensity of provocation and emotional displays as punishment outcomes to change the visual representation of opponent faces. Results will be integrated and summarized within a cybernetic framework describing the appraisal process and subsequent action selection.

CHAPTER

II

Chapter 2 - Facing the Enemy

Empirical Synopsis

The suffering of an opponent is an important social affective cue that modulates how aggressive interactions progress. To investigate the affective consequences of opponent suffering on a revenge seeking individual, two experiments (total N = 82) recorded facial muscle activity while participants observed the reaction of a provoking opponent to a (retaliatory) sound punishment in a laboratory aggression task. Opponents reacted via pre-recorded videos either with facial displays of pain, sadness, or neutrality. Results indicate that participants enjoyed seeing the provocateur suffer: indexed by a coordinated muscle response featuring an increase in zygomaticus major (and orbicularis oculi muscle) activation accompanied by a decrease in corrugator supercilii activation. This positive facial reaction was only shown while a provoking opponent expressed pain. Expressions of sadness, and administration of sound blasts to non-provoking opponents, did not modulate facial activity. Overall, the results suggest that revenge-seeking individuals enjoy observing the offender suffer, which could represent schadenfreude or satisfaction of having succeeded in the retaliation goal.

2.1 Introduction

Escalating aggressive interactions are often characterized by desires to take revenge and to punish the other person for a previous interpersonal offense (Jackson et al., 2019). While many studies investigated the antecedents of revenge seeking, only few published studies examined reactions to retaliatory actions that would satisfy the individual's need for vengeance. The present research investigated how the affective reactions of avengers are modulated by the target's reaction to a (retaliatory) sound punishment, as indexed by spontaneous muscle activations during the observation of the target's suffering.

Several behavioral and neuropsychological studies indicate that taking revenge can be rewarding for the aggressor (for a review see, Chester, 2017). In one study, for example, participants' brains were scanned using positron emission tomography (PET) during punishment of a defector in an economic exchange. The brain scans revealed increased activation of the *anterior dorsal striatum*—a region that is typically implicated in the anticipation of rewards (de Quervain, 2004). Another neuroimaging study measured BOLD responses during physical punishments of a provocateur in a behavioral aggression paradigm (Chester & DeWall, 2016). This study found greater activity in the *nucleus accumbens* (NAcc), which is a brain region critical for the subjective experience of hedonic reward. Evidence for a pleasant experience of revenge was also found in measurements of event-related potentials in scalp-recorded EEG (Krämer et al., 2008). After feedback that they could punish the opponent in a behavioral aggression paradigm, participants with high trait aggressiveness showed an enhanced negativity at mediofrontal sites in the EEG relative to participants with low trait aggressiveness. This difference suggests that punishment of the opponent was rewarding for the high trait aggressive participants or less rewarding for low trait aggressive persons. In sum, several research findings point to the conclusion that taking revenge could be rewarding for the revenge seeking person.

In the studies reviewed above, participants could not see the target's reaction to the punitive action. Hence, they leave it unclear how specific reactions of the retaliation target affect revenge seeking. This question was examined in a study in which participants could observe the opponent's reaction to physical punishments in a laboratory aggression paradigm (Eder et al., 2020). In this study, participants played a competitive reaction time (RT) game against (fictitious) opponents who were supposedly connected via internet. The winner in the game could administer an annoying sound blast to the opponent, whose intensity from mild to severe was selected before each trial. Some opponents showed a very aggressive behavior during the game by consistent administrations of severe sound blasts to the participant, which provoked retaliatory sound punishments from the participant. Importantly, participants saw a video transmission of the opponent in winning trials, in which he expressed pain, anger, sadness, or no emotion in reaction to the sound punishment. A meta-analysis of four experiments showed that seeing the opponent in pain reduced intensities of subsequent noise punishments most strongly (r = .39, [.27, .51]), while displays of sadness (r = .16, [-.05, .36]) and anger (r = .02, [-.12, .16]) had no appeasing effects.

Several explanations can be proposed to understand why individuals reduced revenge-seeking in the study of Eder et al. (2020) after they have seen the provocateur suffer. One explanation is that provoked participants wanted to make the opponent pay for their prior provocation, and that seeing the provocateur suffer satisfied this need (Frijda, 1994). According to this *satisfaction hypothesis*, participants should be pleased about seeing the opponent suffer from a retaliatory punishment. Alternatively, it is also possible that participants were emotionally distressed by the observation that they have visibly harmed another person, for instance, by instigation of guilt feelings (Baumeister et al., 1994) or by empathetic concerns for the opponent (Young et al., 2017). According to this *empathetic-concern/guilt hypothesis*, the view of a suffering opponent should evoke an unpleasant experiential reaction in the observer. In sum, affective reactions of opposite valence can be hypothesized for seeing the provoking opponent hurt, which were examined in the present research using facial electromyography (fEMG).

The Present Research

The present studies were designed to measure affective reactions towards revenge punishment outcomes in a structured competitive game. Participant and (fictitious) opponents interacted with each other via a competitive RT game similar to that used by Eder et al. (2020). During each game round, the participant was asked to press the mouse button based on a visual cue. The main goal was to act faster as the opponent, since whoever presses the button last during each round will be subjected to an aversive sound blast via headphones. Participants were instructed to select the intensity of the sound blast that their opponent would be subjected to in case the opponent lost the game round. If participants themselves lost the game, they received a sound blast that was selected by the opponent. Most opponents selected low intensities, but a few of them consistently selected high intensity sound blasts to provoke retaliatory punishments by the participants. Only those participants who punished provoking opponents more strongly compared to non-provoking ones were included in the analyses according to a pre-registered selection rule (https://osf.io/74gh5/).

Most important, participants could observe the (male) opponent during the sound punishment via an allegedly live video feed (without audio). In most of the trials, the opponent showed a calm (neutral) face to the sound punishment; however, in selected trials he reacted with a clear facial expression of suffering (pain).

Participants' own facial activities during viewing were recorded via electrodes positioned on muscles in the face.

Facial EMG is a highly time sensitive and unobtrusive measurement procedure of muscle contractions. It can register even subtle shifts in muscle tone that are spontaneously produced during the viewing of emotional stimuli such as affective pictures or videos (Cacioppo et al., 1986; Jäncke, 1994). Furthermore, fEMG allows for a measurement of affective reactions without interruption of the revenge act and drawing attention to the measurement of affective reactions. The latter could be important because there are strong normative rules in respect to inflicting pain on others which could distort self-reports (Vigil-Colet et al., 2012).

To differentiate between negative and positive affective reactions, we recorded muscle activations of the *corrugator supercilii* (CS, brow furrowing) and the *zygomaticus major* (ZM, lifting mouth corners, smiling) that index unpleasant and pleasant experiential reactions, respectively (Jäncke, 1994). In addition, a coordinated increase in both ZM and CS activity during the observation of a pain expression could be interpreted as a simulation of pain via facial mimicry (Sun et al., 2015).

In Study 1, opponents reacted to a sound punishment either with expressions of pain or with no emotion (calm face). In Study 2, an additional condition was included in which opponents reacted with sadness. Sadness was included in Study 2 because it is an expression of a negative emotion that is not a specific reaction to physical pain. By comparison, it could hence be determined whether facial activities are specific to expressions of physical pain.

2.2 Study 1

Study 1 investigated muscle contractions of CS and ZM during the observation of opponents that expressed pain or no emotion after a (retaliatory) sound punishment. According to the comparative-suffering hypothesis, seeing the opponent hurt by the retaliatory sound punishment should be appraised positively by the avenger, as indexed by an increased ZM activation. According to the empathetic-concern/guilt hypothesis, by contrast, seeing the opponent suffer from a sound punishment should increase CS activity, due to a personal distress reaction, or increase both, ZM and CS activities, in line with an empathic reaction involving pain mimicry.

2.2.1 Method

Preregistration documents, materials, experiment files, and raw data can be accessed at https://osf.io/74gh5/.

Participants

A total of 56 volunteers were recruited from the participant pool of the University of Würzburg. Six participants were excluded due to equipment failure and an additional four participants due to excessive movement artifacts. In line with our pre-registered exclusion criteria, an additional six participants were excluded because they showed no retaliatory behavior (i.e., provoking opponents were not more punished than non-provoking opponents). The final sample consisted of 40 participants (5 male, *M* age = 23.55 years, *SD* = 3.69). A sensitivity analysis showed that this sample size had sufficient statistical power $P = .80$ for the detection of an ANOVA effect $f \geq .20$ of the opponent's emotional reaction on the participant's facial activity (correlation among measures: $r = .20$, nonsphericity correction = 1, performed with GPower 3.1.9.2). Males and females were recruited because gender differences are negligible after provocation (Bettencourt & Miller, 1996). All participants gave prior informed consent and they received 15 € for participation. The study was performed in agreement with the Declaration of Helsinki and approved by the ethics committee of the University of Wuerzburg (GZEK 2020-74).

Materials

The opponent reaction videos were taken from Eder and colleagues (2020). The videos were 3000 ms long and showed young males wearing headphones. Only young males were selected as targets of aggression to control for target effects and gender differences in the expression of pain (Wise et al., 2002).

Suffering was expressed with facial displays of pain involving brow furrowing, teeth clenching, and a rapid shutting of the eyes (see the video material at https://osf.io/ysnd3/). In the videos with no emotion expression, the opponent showed no visible reaction to the sound punishment. We included a total of 24 videos displaying pain reactions (two for each model). Emotional ratings of the video material by an independent sample (N = 289) are provided in Appendix A (Table S5 & S6). Additionally, we included 24 videos where the opponent displayed smiling reactions. These videos were only shown during noise punishments in a subset of trials in which the participant has lost the game and received a punishment by the opponent. Smiling responses were included to examine how the receipt of a noise punishment influences facial mimicry. Given that their analyses address another research question, they are not reported in this paper.

For noise punishment, a 3 second long white noise was used. Noise blasts were taken from the Inquisit database (Millisecond Software, Seattle, WA, USA). The loudest noise blast (5) was 75 dB and intensities were lowered in 5dB steps corresponding with each noise level step (1= 55 dB, 5= 75 dB).

Participants also answered the German version of the Interpersonal Reactivity Index (Davis, 1983), the so called SPF (Paulus, 2009), a self-report measure of dispositional empathy that is positively related to facial mimicry effects (Drimalla et al., 2019; Sonnby-Borgström et al., 2003). The SPF comprises four subscales (empathic concern, perspective taking, fantasy scale and personal distress) with 4 items each.

Procedure

The skin was prepared using alcohol and an abrasive electrolyte solution before placing the Ag/AgCl surface electrodes (4mm) on top of the ZM muscle, the CS muscle, and on the left mastoid serving as a reference (electrode distance: 1.5 cm). Electrode handling and placement were conducted in line with the guidelines for human electromyographic research (Fridlund & Cacioppo, 1986). EMG data were recorded with a 16-channel amplifier (V-Amp, Brain Products, Gilching Germany). The raw signal was stored on a separate computer.

Written instructions informed participants that they would play a competitive RT game (CRTT) against an opponent who would be visible via a live video feed. Participants were instructed that the one who reacts the slowest would be punished via an unpleasant noise blast. The task procedure was a modified Taylor aggression paradigm (Taylor, 1967). Before the CRTT began, the noise blasts were played in the highest and lowest setting to familiarize the participant with the sound levels. A cover story told participants that skin conductance levels would be measured with electrodes as an indicator of stress experienced during the task. Instructions also stated that their opponents would be students located at a different university in Germany and that they were supposedly assigned to a study condition in which they could observe the opponent via internet video transmissions, whereas the opponent could not see them. The task was run using E-Prime 3.0 software (Psychology Software Tools, Sharpsburg, PA, USA) on a computer with a 1920 x 1200 screen for stimulus presentation. Responses were collected via mouse clicks.

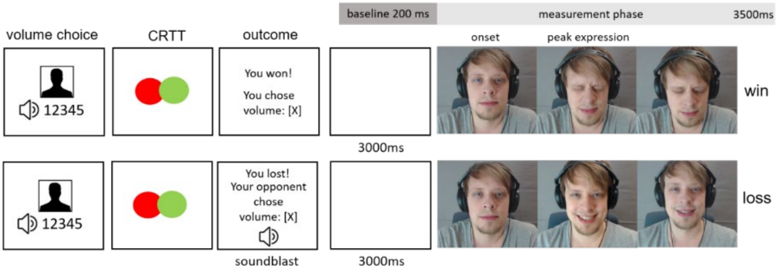
Figure 3 shows the sequence of events in a trial of the CRTT. Each trial began with the selection of the level of the noise blast (1-5) that would be administered to the opponent if the participant won the trial. The selection was followed by the reaction time task, which showed a red circle as a preparatory signal that changed into green as a go signal for a rapid click of the left mouse button. The time window for a valid response was set to 1000ms. An error message appeared if the participants did not press the key during this time limit ("We could not detect your response inside the time window. Please repeat this trial."). After a valid response the participant was informed about the outcome of the reaction task (win/loss). In win trials, the following message was displayed for 3000ms: "You won. You chose volume [value]. You will now see your opponent as he hears the sound." After a 3000ms blank display, the opponent video was shown (3000ms). In loss trials, the participant received the message: "You lost. Your opponent chose volume [value]." They heard the noise blast of the indicated intensity, and after 3000ms blank they watched a video of the opponent. The opponent showed either a smile or no emotion when he won the game. Participants' facial activities during his reaction were recorded but not analyzed for this paper.

Overall, participants played 48 blocks, 16 in which the fictitious opponent provoked retaliatory punishments with administrations of very loud sound blasts (levels 4-5) and 32 blocks in which opponents did not provoke with selections of mild sound blasts (levels 1-3). Each block consisted of 5 CRTT trials. In 16 blocks with non-provoking opponents, the outcome of the game (win/loss) was random. In the remaining 32 blocks, the outcome in the first trial was randomly determined, the second and third trials were lost (the participant received sound blasts from the opponent), and the fourth and fifth trials were won by the participant (administration of sound blasts to the opponent). When the provoking opponent lost the game in the fourth and fifth trials, he expressed pain in the fourth trial and no emotion in the fifth trial. In four of the 16 blocks, however, he expressed no emotion in both trials. The remaining 16 blocks with non-provoking opponents had analogous win/loss streaks and opponents displayed analogous reactions to sound punishments. Each opponent was featured four times (total number of opponents = 12).

After each block, participants were asked to rate their feelings of anger towards the opponent on a 5 point Likert scale and feelings of dominance, arousal and pleasantness on self-assessment manikin scales (Bradley & Lang, 1994). After the CRTT, participants answered the SPF questionnaire and were debriefed. The debriefing stated that all opponents were fictitious and therefore each received punishments was controlled by the experimental software and that all reactions were prerecorded videos.

Figure 3

Example of a win and loss trial featuring video reactions.



Data Preprocessing and Analysis Plan

EMG data were processed offline using the Brain Vision Analyzer 2.2 software (Brain Products, Gilching Germany). fEMG data were filtered (20 Hz low cutoff filter, 499 Hz high cutoff filter, 50 Hz notch), full wave rectified and segmented (-200 ms prior to stimulus onset to 3500ms post stimulus onset). Prior to data analysis, trials were averaged for each condition and exported into 500ms time bins. Data were additionally screened for low EMG activation (caused by technical failures such as loose or broken electrodes). Artefacts were semi automatically detected via a build in function of the brain vision analyzer using a gradient of maximum allowed voltage steps of 50 µV/ms and a maximum difference of 200 µV in each 200ms interval. Additionally, data were visually inspected for movement artefacts (e.g., coughing) and all trials containing artefacts were removed. Difference scores were calculated using the 200ms prior to video onset as baseline. All figures display standardized means (within-subject z transformed difference scores) for a more convenient interpretation of the results.

Deviating from our preregistered analysis, we ran a linear mixed model (LMM) analysis in JAMOVI (using the GAMLj module, version 1.0.7; Gallucci, 2019), upon reviewer suggestion. We included the difference score of the EMG response (baseline corrected) as the dependent variable and added fixed effects of *Provocation* (high / low), *Emotion* (neutral / pain), *Time* (500 ms intervals over 3500ms), as well as interactions thereof. We included participant as a random effect. Significance was calculated via the Satterthwaite's method to estimate degrees of freedom and generate p-values. The model specification was as follows: MuscleResponse ~ 1 + Provocation + Emotion + Time + Provocation:Emotion + Provocation:Time + Emotion:Time+(1 | Subject).

Since our main hypotheses are only concerned with muscle reactions during high provocation blocks, we followed up significant *Provocation* by *Emotion* interactions with simple effects analysis, with provocation as a moderator. This approach allows us to compare muscle responses between the negative emotion conditions (pain / sadness) with the neutral condition separately for each provocation condition. We further conducted simple effects analysis of time with emotion as a moderator, separately for each provocation condition. We will only report and discuss the results from each main effect, the interaction between provocation and emotion as well as our a priori planned comparisons of muscle responses based on opponent expressions within each provocation condition (simple effects for emotion and time, for each provocation condition separately). For a report on all fixed effects results including each possible interaction term please see Appendix A.

2.2.2 Results

Ratings

As shown in Table 1, after provoking compared to non-provoking blocks participants were more aroused, $t(39) = 4.28$, $p < .001$, $d_z = 0.68$, felt less dominant, $t(39) = 3.50$, $p = .001$, $d_z = 0.55$, and more unpleasant, $t(39) = 4.98$, $p < .001$, $d_z = 0.79$. Selected levels of noise punishments were significantly higher after provocation, $t(39) = 8.38$, $p < .001$, $d_z = 1.33$. Self-reported anger was also higher $t(39) = 9.13$, $p < .001$, $d_z = 1.44$.

Punishment Choices

Changes in revenge seeking (volume choices) were indexed by difference scores that compared volume choices in CRTT trials before and after having seen an opponent reaction. Difference scores were calculated via subtracting the chosen volume in Trial 4 (win CRRT trial with emotional opponent reaction) from the chosen volume in Trial 5 (after having observed the opponent video). Negative values indicate a reduction in the intensity of desired noise punishments. Results showed that provoked participants significantly reduced the volume punishment following pain displays (M = -0.37, SD = 0.50) compared to neutral displays (M = -0.14, SD = 0.59), $t(39) = -2.41$, $p = .021$, $d_z = 0.38$; in contrast, there was no significant effect without prior provocation, $t(39) = 1.40$, $p = .171$, $d_z = 0.22$.

Table 1

Means (with SD) for behavioral measurements in Study 1

	Noise Level	Valence	Arousal	Dominance	Anger
provocation	2.99 (1.13)	3.63 (0.71)	2.12 (0.80)	3.51 (1.09)	3.07 (1.04)
no provocation	2.00 (0.82)	3.82 (0.69)	1.80 (0.65)	3.71 (1.09)	1.94 (0.68)

Note. Ratings are based on 5-point scales with 1 indicating no/the least amount and 5 indicating the largest amount, except for valence which ranged from 1 negative to 5 positive.

Electromyography

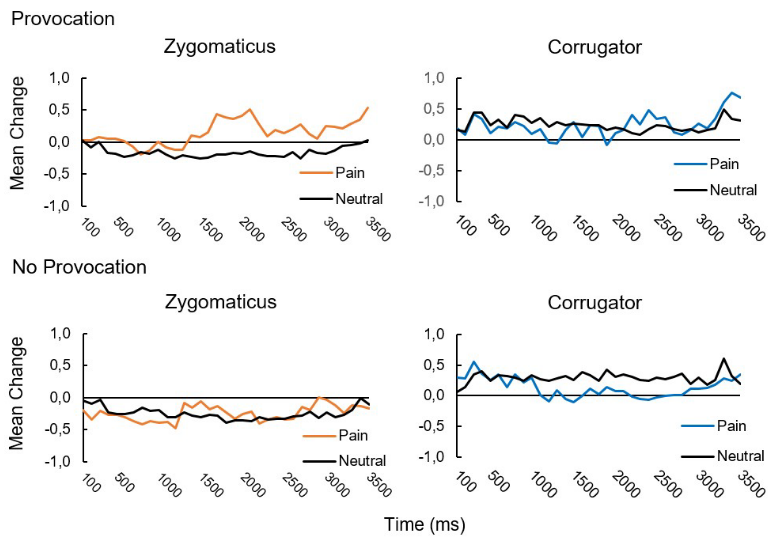
Zygomaticus Major (ZM)

Fixed effect omnibus tests of ZM muscle responses resulted in significant main effects of provocation, $F(1,1059) = 27.66$, $p < .001$ and emotion, $F(1,1059) = 12.79$, $p < .001$, along with a significant provocation by emotion interaction, $F(1,1059) = 16.53$, $p < .001$. The results indicate that participants' ZM muscle responses depended on the emotional expressions of the opponent and that this dependency is further affected by provocation. Simple effects analysis with provocation as a moderator indicate that ZM muscle responses following neutral expressions were significantly lower during high provocation trials when compared to responses following pain expressions ($b = -31.55$, $SE = 5.84$, $t = -5.40$, $p < .001$). As shown in Figure 4, viewing the provoking opponent suffer from the noise punishment led to an increase in ZM muscle activity compared to neutral opponent reactions. The ZM reaction started about 1300s after video onset. We obtained no difference in ZM muscle reactions during the low provocation condition ($b = 2.02$, $SE = 5.84$, $t = 0.35$, $p = 0.729$). Analysis of simple effects of time within each emotion, separately for each provocation condition, did not result in any significant effects of time, all Fs ≤ 0.87, all ps $\geq .541$.

Corrugator Supercilii (CS)

Fixed effect omnibus tests of CS muscle responses resulted in significant main effects of provocation, $F(1,1059) = 6.95$, $p = .009$, and a significant provocation by emotion interaction, $F(1,1059) = 5.69$, $p = .017$. We obtained no significant main effect of emotion, $F(1,1059) = 0.269$, $p = .604$. The absence of a clear main effect of emotion and the rather weak interaction of provocation and emotion indicate that CS responses were largely affected by provocation irrespective of the opponents' emotional expression. Simple effects analysis with provocation as a moderator indicate lower CS responses following neutral expressions during high provocation when compared to responses following pain expressions (b = -33.4, SE = -16.2, t = 2.05, p = 0.040). We obtained no difference in CS muscle reactions during the low provocation condition (b = 21.4, SE = 16.2, t = 1.32, p = 0.187). Analysis of simple effects of time within each emotion, separately for each provocation condition, did not result in significant effects, all Fs ≤ 1.08, all ps ≥ .374.

Figure 4



Note. Time course of the normalized EMG response from stimulus onset to 3500 ms post-stimulus onset during observation of suffering opponents in Study 1.

2.2.3 Interim Discussion

Results show an increase in ZM muscle activation during observation of a pain response. Assuming that the ZM is responsive to positive affects (Brown & Schwartz, 1980), this increase indicates that avengers enjoyed seeing the opponent suffer from the noise punishment. The effect was only observed after punishments of provoking opponents, which suggests that revenge motivation played an important role in the appraisal of opponent reactions.

Furthermore, ZM activity did not change after the punishment of provocateurs who did not express suffering. This result pattern suggests that participants were specifically pleased about having hurt the opponent. Analyses of CS activity did result in a higher level of CS activity within the high provocation condition. However, this effect was considerable smaller than the ZM activation and does not entail evidence for pain mimicry (activations in ZM and CS did not match timewise) or modulations instigated by feelings of guilt, as indexed by increased CS activity in the absence of ZM activation. Observation of a suffering non-provoking opponent did also not trigger an empathetic CS reaction, which is in line with previous studies showing that empathetic concerns are reduced in competitive task settings (Zaki, 2014). In short, facial activities supported the satisfaction hypothesis and were not in line with the empathetic-concern/guilt hypothesis.

2.3 Study 2

Study 2 included an additional opponent reaction to sound punishments: sadness. Sadness expresses suffering that is not specific for physical pain (Horstmann, 2003). In the context of the competitive RT game, the expression could be also interpreted as a signal of defeat and/or submission (Tiedens, 2001). By comparing pain with an additional suffering display, we aimed to investigate whether the ZM increase observed in Study 1 is specific for expressions of pain. In Study 2, we additionally measured activations of the *orbicularis oculi* (OO) muscle surrounding the eye. In combination with an increase in ZM activity, OO activity is a signifier of 'genuine' smiling (the so-called Duchenne smile; Ekman & Friesen, 1982), see also Hess et al. (2017).

2.3.1 Method

Participants

Seventy-six students participated for a monetary payment of 12€. In line with our exclusion criteria, 18 datasets were excluded due to artefacts and technical errors (caused by broken electrodes and amplifier malfunction) and an additional 16 datasets due to ineffectiveness of the provocation treatment. The final sample consisted of 42 participants (7 males, mean age = 25.07 years, $SD = 4.43$). A sensitivity analysis showed that this sample size had sufficient statistical power $P = .80$ for the detection of a small ANOVA effect $f \geq 0.20$ of the opponent's emotional reaction on the participant's facial activity (correlation among measures: $r = .20$, nonsphericity correction = 1, performed with GPower 3.1.9.2). All participants gave prior informed consent.

Materials

We selected 12 pain videos from Study 1 and added 12 sadness videos taken from Eder et al. (2020). Emotional ratings of the video materials are provided in Appendix A. For the registration of prolonged or late facial reactions, a still image taken from the last video frame was appended to the video for a duration of 3 seconds. The still image depicted the negative emotional expression featured in the video and prolonged the expression unobtrusively.

Procedure

To measure the *orbicularis oculi* muscle activation, we placed two electrodes next to the outer corner of the left eye. All other EMG specific procedures were identical to those of Study 1. To account for later changes in muscle activation following the peak expression of each emotion, we prolonged the time window of the EMG measurement from 3 to 6 s.

The CRTT, self-assessment manikin scales, anger item and the SPF empathy questionnaire were identical to Study 1. Overall, there were 12 blocks: 6 with provoking and 6 with non-provoking (fictitious) opponents that appeared in random order. The number of trials per block was increased to 10 trials. The win/loss ratio was 5/5. Opponents' reactions were only shown during punishments in winning trials. In two out of five winning trials, the opponent displayed pain or sadness as a reaction to the sound punishment. In the other three out of five winning trials, the opponent expressed no emotion. Blocks always featured either a pain or a sadness expression twice, but never a combination of both. In total, this procedure yielded 6 pain and 6 sadness expressions by provoking and non-provoking opponents, respectively. Participants played two blocks against each opponent:

in one block the opponent reacted with sadness, in the other block with pain. In contrast to Study 1, opponent videos were only shown during win trials.

Data preprocessing and Analysis Plan

Data processing and artefact detection were identical to Study 1. Due to the longer recording period, we segmented data from 500 ms prior to stimulus onset to 6000ms post stimulus onset into 500 ms time intervals. Difference scores were calculated using the 500 ms time bin prior to stimulus onset as baseline.

We included the difference score of the EMG response (baseline corrected) as the dependent variable and added fixed effects of *Provocation* (high / low), *Emotion* (neutral / pain / sadness), *Time* (500 ms intervals over 6 seconds), as well as interactions thereof. We included participant as a random effect. All analysis strategies and the LMM approaches are identical to those of Study 1. Figure 5 shows an overview of standardized muscle activations relative to baseline in each condition.

2.3.2 Results

Ratings

Table 2 provides a descriptive overview of the rating measures. In blocks with provoking opponents, participants felt more aroused, $t(41) = 5.05$, $p < .001$, $d_z = 0.75$, and less dominant, $t(41) = -3.16$, $p = .003$, $d_z = 0.43$; however, there was no significant difference in ratings of pleasantness, $t(41) = -1.19$, $p = .234$,

$d_z = 0.18$. Participants were more angry about provoking compared to non-provoking opponents, $t(41) = 9.51, p < .001, d_z = 1.46$. In line with the pre-registered manipulation check of the provocation treatment, provoking opponents were punished more than non-provoking ones, $t(41) = 8.21, p < .001, d_z = 1.27$.

Table 2

Means (with SD) for measurements in Study 2

	Noise Level	Valence	Arousal	Dominance	Anger
provocation	3.04 (0.98)	3.83 (0.63)	2.13 (0.74)	3.54 (0.97)	2.82 (0.91)
no provocation	1.89 (0.87)	3.86 (0.55)	1.67 (0.70)	3.80 (0.87)	1.43 (0.58)

Note. Ratings are based on 5-point scales with 1 indicating no/the least amount and 5 indicating the largest amount, except for valence which ranged from 1 negative to 5 positive.

Punishment Choices

Difference scores for the emotion condition were calculated by subtracting the noise level selected for the trial with an emotional opponent reaction from the noise level selected in the subsequent trial. Analyses of difference scores showed that provoked participants significantly reduced the volume of the noise punishment following pain displays (*M* = -0.44, *SD* = 0.81) compared to neutral displays (*M* = -0.09, *SD* = 0.63), $t(41) = -2.26$, $p = .029$, $d_z = .349$. Sadness displays did not result in a significant decrease in punishment volume compared to neutral displays (*M* = -0.17, *SD* = 0.42), $t(41) = -0.381$, $p = .705$, $d_z = .058$. Difference scores for neutral displays were calculated based on the last two neutral trials (before any emotional reaction occurred) for each block separately (pain/sadness). Analogous comparisons of difference scores in no provocation blocks produced no significant effects, all *t*s(41) ≤ 0.82, all *p*s ≥ .418.

Zygomaticus Major (ZM)

Fixed effect omnibus tests of ZM muscle responses resulted in a significant main effect of Provocation, $F(1,2933) = 41.20$, $p < .001$, a significant main effect of Emotion, $F(1,2933) = 3.31$, $p = .037$, and a significant Provocation by Emotion interaction, $F(1,2933) = 13.60$, $p < .001$. This pattern of effects indicates that ZM muscle reactions were depended on the opponents' emotional expressions as well as the level of provocation.

ZM activity after high provocation. ZM activations following expressions of pain were significantly higher than ZM activations following neutral expressions ($b = 136.2$, $SE = 26.6$, $t = 5.13$, $p < .001$). Sadness expressions also resulted in a significantly higher activation of ZM when compared to neutral ($b = 71.6$, $SE = 26.6$, $t = 2.70$, $p = .007$). The analysis of time effects revealed a significant increase in ZM activity following pain during time bins 9-12 (all *b*s ≥ 148.18, all *p*s ≤ .05), as well a significant increase following sadness during time bins 10-12 (all *b*s ≥ 195.97, all *p*s ≤ .01).

ZM activity after low provocation. ZM activity following pain expressions was significantly lower compared to neutral ($b = -54.7$, $SE = 26.6$, $t = -2.06$, $p = .040$). Responses following sadness expressions did not differ from neutral ($b = 14.1$, $SE = 26.6$, $t = 0.531$, $p = 0.596$). The analysis of time effects revealed no significant effects of time, all *F*s ≤ 0.92, all *p*s ≥ .518.

Orbicularis Oculi (OO)

Fixed effect omnibus tests of the OO responses revealed a main effect of Provocation, $F(1, 2933) = 14.96$, $p < .001$, Emotion, $F(2,2933) = 12.22$, $p < .001$, and a significant interaction between Provocation and Emotion, $F(2, 2933) = 5.20$, $p = .006$. This indicates that OO responses were dependent on provocation as well as emotional opponent expressions.

OO activity after high provocation. OO activations following pain were significantly larger compared to neutral ($b = 156.7$, $SE = 27.0$, $t = 5.03$, $p < .001$). OO activations following sadness expressions did not differ from neutral ($b = 24.9$, $SE = 31.1$, $t = 0.799$, $p = .425$). The time specific analysis revealed a significant increase in OO activity during time bins 7-12 following pain expressions (all *b*s ≥ 252.94, all *p*s ≤ .004). We also observed a significant decrease following sadness expressions during time bins

4-7 (all *b*s ≥ -201.87, all *p*s ≤ .022) followed by a significant increase during time bins 9, 10 and 12 (all *b*s ≥ 191.28, all *p*s ≤ .030). The initial decrease in OO activity was also present following neutral expressions during time bins 3-5 (all *b*s ≥ -191.39, all *p*s ≤ .030).

OO activity after low provocation. OO responses following pain expressions were did not significantly differ from neutral (*b* = 60.00, *SE* = 31.1, *t* = 1.92, *p* = .054). Responses following sadness expressions were significantly higher compared to neutral (*b* = 66.3, *SE* = 31.1, *t* = 2.131, *p* = .033). We also observed an early significant decrease of OO activity following neutral expressions during time bins 2-4 (all *b*s ≥ -173.52, all *p*s ≤ .05), as well as following sadness expressions during time bins 2-4 and 6-7 (all *b*s ≥ -225.55, all *p*s ≤ .010). There was only a short significant decrease during time bin 2 (*b* =-221.29, *p* = 0.012) following pain expressions (all other *b*s ≥ 160.22, all *p*s ≤ .069).

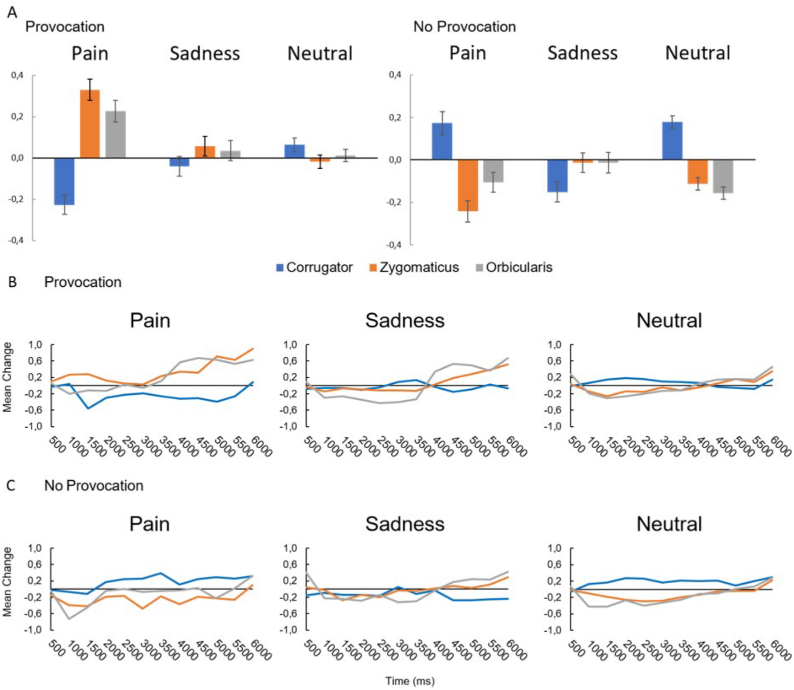
Corrugator Supercilii (CS)

Fixed effect omnibus tests of CS responses revealed no significant main effect of Provocation *F*(1, 2933) = 2.01, *p* = .156, but a significant effect of Emotion, *F*(2,2933) = 6.32, *p* = .002, and a significant interaction between Provocation and Emotion, *F*(2, 2933) = 37.31, *p* < .001. This indicates that CS responses were dependent on opponent expressions and this dependency was affected by provocation.

CS activity after high provocation. CS activity following pain expressions was significantly lower compared to neutral (b = -213.2, SE = 40.2, t = 5.306, $p < 0.001$). We observed no significant difference between sadness and neutral (b =20.8, SE = 40.2, t = 0.517, p = .606). We further observed no significant effects of time bs ≥ 176.12, all ps ≤ .12).

CS activity after low provocation. Responses following pain did not differ significantly from neutral expressions (b = 50.6, SE = 40.2, t = 1.26, p = .208). Responses following sadness expressions were significantly lower compared to neutral (b = -205.8, SE = 40.2, t = 5.12, $p < .001$). We observed no significant effects of time (all bs ≥162.419, all ps ≤ .153).

Figure 5



Note. Panel A represents the z-standardized means of muscle activation over the full 6s period for each emotion expression and muscle site (CS, ZM, and OO) separately. Panel B and C depict the time course of the z-transformed fEMG response from stimulus onset to 6000ms post-stimulus onset for each provocation and emotion condition.

2.3.3 Interim Discussion

Seeing the provoking opponent suffer from a (retaliatory) sound punishment increased smiling indexed by ZM and OO activity, while CS activity decreased, which--taken together--suggests a “genuine” smiling reaction. The smiling reaction was only evoked by displays of pain, whereas the opponent’s expression of sadness had no analogous effect. This pattern suggests that the smiling reaction was specific for displays of pain as an expression of suffering.

2.4 General Discussion

Two experiments recorded participants’ facial muscle activities while they observed the emotional reaction of an opponent to a (retaliatory) sound punishment in a laboratory aggression task. The fEMG data revealed that participants enjoyed seeing the provoking opponent suffer from the noise punishment: there was a coordinated increase in ZM (and OO muscle in Study 2) activation that was accompanied by a decrease in CS activation in Study 2. The smiling reaction only occurred during observation of punishments of provoking opponents, and participants did not smile when the provoking opponent expressed no emotion or sadness. The smiling during the opponent’s expression of pain, but not during expressions of sadness in Study 2, suggests that the enjoyment was specifically linked to the perception that the sound punishment has physically harmed the opponent.

Participants' smiling after having visibly hurt the provoking opponent could reflect satisfaction of having succeeded in the retaliation goal to inflict harm on this opponent. This would fit previous research that also found evidence for aggressive pleasure and rewarding effects of revenge taking (for a review see Chester, 2017).

Besides satisfaction, participants could also have experienced schadenfreude over the opponent's misfortune, which is known to produce a smiling reaction that is indistinguishable from joy (Boecker et al., 2015). Hence, several explanations are possible for why participants smiled during painful punishments of the provocateur, and more research is needed to distinguish between them. One difference between Schadenfreude and aggressive pleasure is agency. While Schadenfreude is associated with the misfortune of others, without any personal responsibility or agency, reward and pleasure stemming from reaching the retaliation goal should increase with agency. Future studies could vary the amount of subjective agency while administering the punishment and therefore differentiate between the affect evoked by carrying out the aggressive act itself and the affect evoked by the punishment outcome (for example suffering).

Perception of the opponent's pain (or sadness) did not trigger mimicry and/or a compassionate facial reaction, as indexed by increased activity of the CS muscle. The absence of a compassionate response to the suffering of the provocateur is in line with reports that empathy for pain is markedly reduced for unfair and disliked persons (Likowski et al., 2008; Singer et al., 2006). However, participants also showed no compassionate reaction towards non-provoking opponents, who were presumably more liked in comparison. Research suggested that facial mimicry is generally decreased in competitive task settings (Weyers et al., 2009), which could explain the absence of a mimicry response.

The present data also do not indicate that participants experienced guilt due to having visibly harmed the opponent. For this discussion, however, it should be noted that participants were only included in the present analyses when the provocation of retaliatory aggression by opponents was effective. It is possible that participants who do not seek revenge show more compassion with targets of sound punishments, which could be an interesting avenue for future research.

Analyses of punishment choices replicated the results from a previous study that observed a reduction of punishment after observation of a suffering response (Eder et al., 2020). Like in Eder et al.'s study, noise punishments were significantly reduced after observation of pain but not after sadness and calm displays. The present study additionally showed that this effect critically depends on a prior provocation, which is a novel and original finding. [In Eder et al.'s study, only provoking opponents reacted emotionally to noise punishments]. An interesting question for future research is whether the experience of positive affects (as indexed by ZM activities) is linked to a down regulation of punishment (indexed by a reduction of punishments intensities).

The design of the present studies was not optimized to assess this relation. Future studies could investigate correlations between the magnitude of smiling after having harmed the opponent and subsequent revenge seeking, and whether the experience of positive affect is causally involved in the seeking revenge.

Several limitations of the present study need to be mentioned. First, regarding the stimuli, full expression onsets and overall intensity of expressions varied slightly across the different opponents. This adds variance to existing individual variations in reaction onsets. Further, differences in the intensity of the portrayal of pain and sadness expressions can arouse observers in different degrees, which in turn could modulate facial activities (Fujimura et al., 2010).

Secondly, participants were mainly females playing against exclusively male opponents. Previous studies demonstrated gender differences in display rules as well as in the perception of pain cues displayed by men and women (Decety et al., 2008; Nayak et al., 2000). These studies mainly focused on either beliefs about pain tolerance or the empathy evoking aspects of pain perception. Studies which investigated pleasure evoked by observing disliked others who experience misfortunate events, report no gender differences in overall pleasure (Hareli & Weiner, 2002). Further, studies focusing on schadenfreude and facial muscle reactions in men (Boecker et al., 2015) reported the same muscle response pattern (ZM and OO increase, CS decrease) we observed. However, given that our sample consisted of mainly female participants, it is important to corroborate our findings with further research, that takes both, opponent and participant gender, into consideration.

Thirdly, not every participant showed a positive affective response towards opponents in pain. Although the majority of participants exhibited a smiling response, visual analysis of muscle activation patterns on individual levels also revealed a reversed pattern for some participants (exhibiting an increase in CS activity towards suffering opponents). The sample of participants showing this reversed reaction was too small for a meaningful analysis (Study 1: n = 3; Study 2: n = 2). Future studies should therefore also take personality traits into account that are known to influence the enjoyment of revenge taking and/or empathetic reactions (e.g., attitudes towards vengeance). Studies also found that victims of an interpersonal offense want offenders to know the reasons for a retaliatory punishment and that they feel most satisfied when the target signals 'understanding' on the victim's intent to punish and a positive moral change in respect to the wrongdoing (Funk et al., 2014; Gollwitzer et al., 2011). Future research could employ a combination of facial displays and messages of understanding as a test for the impact of moral motivations that may also satisfy the avenger.

To summarize, this study demonstrates that avengers enjoy the punishment of offenders if the punishment has hurtful consequences. Further, avengers only adjust their punishment following directly observed hurtful consequences. These hurtful consequences were observed via displays of pain as a salient indicator of opponent suffering. Future studies are needed to distinguish between the different underlying sources of enjoyment, such as schadenfreude or satisfaction with having succeeded in the goal to take revenge.

CHAPTER

III

Chapter 3 - Smiling with the Enemy

Empirical Synopsis

Schadenfreude smiles communicate malicious intentions and are generally frowned upon on a moral level. During aggressive competitions, displays of schadenfreude can potentially increase aggressive punishments, therefore escalating aggressive interactions. In a serious of studies, we measured facial muscle reactions (over the zygomaticus major and corrugator supercilii) towards smiling opponents that were either highly provocative or not provocative, as well as their subsequent punishment, and the categorization of smiles in competitive-aggressive contexts.

Previous research provided only inconsistent evidence concerning the modulation of smile mimicry towards disliked others. In the realm of this study, we tested two competing hypotheses. One being the occurrence of counter empathic responses, constituting of an affective reversal of mimicking behavior, indexed via frowning as a response to smiling opponents. The other being an attenuation of smile mimicry, indexed via a reduction in the smiling pattern towards highly provoking opponents, compared to non-provoking opponents. A significant escalation of punishment when being confronted with Schadenfreude was observed.

The observation of opponent smiles led to imitation behavior (facial mimicry), which was slightly attenuated by previous provocation. The potential functions of smile mimicry in the context of aggressive competition are discussed as containing simulation aspects (to aid in opponent understanding) and the potential mirroring of dominance gestures, to avoid submission.

3.1 Introduction

Smiling is a socially complex facial expression. Smiles serve as tools of communication, ranging from the obvious portrayal of joy to superiority and dominance to the most sinister forms of smiling, malicious joy or schadenfreude (Martin et al., 2017). The communicative functions of smiling are as multifaceted as the range of emotional states smiles are accompanied by (Szameitat et al., 2009), such as smiles masking sadness (Ansfield, 2007). In two studies we investigated how malicious smiling during the delivery of a physical punishment in a competitive task influences spontaneous facial (mimicry) reactions of the punished person viewing the smile, their subsequent retaliatory behavior, and their interpretation of the opponent's smiling intent.

Schadenfreude, defined as the subjective experience of pleasure about another's misfortune (Smith & van Dijk, 2018), has mostly been investigated from the viewpoint of the schadenfreude expressing person, with studies indicating a link between outgroup schadenfreude and the endorsement of potential aggressive acts against the laughed-at outgroups (Shah & Tee, 2019). The process of enabling enjoyment of the negative outcomes for disliked others has been discussed as a reinforcement-based learning process that couples positive affect with the perception of harm, potentially leading to the promotion of violence against disliked groups or individuals (Cikara, 2015). Observing public displays of schadenfreude can foster aggressive tendencies towards the laughed-at person (Lange & Boecker, 2019). Schadenfreude displays are frequent in competitive settings, such as sports (Leach et al., 2003) or when envy is present (Cikara et al., 2014; Smith & van Dijk, 2018).

Being laughed at, as a form of ridicule, is included in many assessments of bullying behavior (Boulton & Hawker, 1997; Einarsen, 2000), being generally regarded as a form of aggressive behavior, either classified as verbal or relational aggression. Since being laughed at generally leads to negative affect in the target, ranging from embarrassment to anger (Platt, 2008), being the target of schadenfreude during an aggressive competition should lead to increases in negative affect and perceived provocation. However, it remains unclear how displays of schadenfreude affect the progression of aggressive episodes.

On the level of face perception, displays of smiling due to schadenfreude are not readily distinguishable from other displays of smiles, since evidence from the measurement of differences in the muscular expression of smile types indicates no difference in the muscular activation patterns observed during joy smiles and schadenfreude (Boecker et al., 2015). Schadenfreude, as a social emotion, may only be correctly interpreted with information about the social context. When being faced with schadenfreude as a response to personal misfortune, the target must decode the facial display as antagonistic via reframing the smiling display due to the social context. Although this is a complex task involving perspective taking and interpreting social cues, perceivers have been shown to be able to differentiate between different types of smiles and their social implications (Rychlowska et al., 2017).

When viewing emotional facial displays, individuals have the automatic tendency to copy the expression of their interaction partner. These so-called *facial mimicry* effects are modulated by interaction context and can serve a communicative function (for review see Hess, 2021). Facial mimicry are influenced by several factors, such as mood (Likowski et al., 2011; Moody et al., 2007) with negative mood resulting in less mimicking tendencies. Negative attitudes towards the other person have been shown to resulting in an attenuation of mimicry responses towards disliked others and outgroups (Likowski et al., 2008). These modulations are also susceptible to changes due to the type of emotion expression to be mimicked (Fischer et al., 2012), with smiles being an expression that is mimicked effortlessly towards liked others, but more effortfully towards neutral or negative others (Blocker & McIntosh, 2016). The observation of smiles, even when devoid of social context, has been shown to activate smiling response tendencies in the perceiver which are difficult to inhibit (Korb et al., 2010), at least towards neutral others.

Engaging in competition affects the magnitude of facial mimicry towards smiles (Weyers et al., 2009b), and has also been shown to change the direction of facial muscle responses--a so called *counter-empathic* response (Lanzetta & Englis, 1989). A counter empathic reaction can be defined as a reversal in valence between the observed expression and the measured response; for example, an opponent smile being responded to with a frown by the perceiver. Lanzetta and colleagues (1989) observed counter empathic mimicking responses towards smiles after participants were confronted with the expectation of a competition against an opponent described as very uncooperative. The authors observed a significantly higher activation of the *depressor anguli oris* (pulling mouth corners down) and *orbicularis oculi muscle* (OO, shutting eyes) in response to smiling opponents. However, the authors did not measure the more commonly investigated muscles to indicate valence, *corrugator supercilii* (CS, brow furrowing) or *zygomaticus major* (ZM, pulling mouth corners up) during their study. Counter empathic facial responses towards smiling have also been reported towards outgroups (van der Schalk et al., 2011). An outward display of schadenfreude during aggressive competition could therefore result in a counter-empathic response since the joy of the opponent is appraised as negative by the perceiver.

Instead of a full reversal of the facial reaction, more recent studies demonstrated an attenuating effect of competition on mimicry responses, observing a weakened form of smile mimicry. In the study of Weyers and colleagues (2009), a competitive mindset was primed with subliminal presentations of competition related words, while measuring facial muscle responses towards static images of smiling or frowning avatars. Results indicated an attenuation of CS muscle relaxation during the observation of smiling competition primed avatars but no differences in ZM activation. No or strongly attenuated facial mimicry responses were also reported when the observer dislikes the smiling person (Blocker & McIntosh, 2016; Likowski et al., 2008). For example, Blocker and colleagues (2016) observed no significant ZM activation nor CS relaxation towards persons smiling that were previously described as aggressive and deceitful.

Schadenfreude smiles occur in contexts where the smile expresses an inherently negative relationship to the recipient. While people are capable of interpreting schadenfreude correctly, it is unclear how the tendency to mimic smiles will interact with the interpretation of smiles as expressing schadenfreude. Research indicates that competitive or aggressive social contexts can attenuate mimicry effects, but these findings rest only on noninteractive passive viewing tasks, not active interactions. A theoretical model that may integrate these findings is the *Mimicry as a Social Regulator* account.

The "Mimicry as a Social Regulator" account of Hess and colleagues (2021) states that mimicry responses can be considered automatic, but their occurrence is modulated via top-down control, based on the social information present and the goals of the perceiver. The goals of mimicking responses should relate to the regulation of social interaction, via strengthening the interaction in responding with mimicking of affiliative displays as long as affiliative motivation is present (Mauersberger & Hess, 2019). In the case of highly negative interactions, such as the aggressive competition in which schadenfreude often takes places, the perceiver should not be motivated to affiliate with the opponent, nor perceive the smiles as a display of affiliation towards her.

To gain insights about the evaluation of schadenfreude smiles and their consequences in aggressive competition, we employed a direct measurement of attribution of smiles via items asking about the attribution of the perceived smiles in a pilot study, as well as an unobtrusive online measure of facial muscle activations during the perception of schadenfreude smiles (Study 1).

Based on the aforementioned empirical evidence, we hypothesized that schadenfreude smiles will result in either counter-empathic responses involving muscle activations indicative of negative affect, or an attenuation of smiling responses, as observed towards disliked others and outgroups in previous studies. We further hypothesized that being laughed at during an aggressive competition should provoke and therefore further escalate the interaction, indicated via an increase of punishment levels following smiling opponent displays.

3.2 Pilot Testing of Smile Interpretations

Since the interpretation of smiles is highly dependent on contextual information, we first conducted a pilot study investigating the interpretation of opponent smiles in a competitive context varying in the level of provocation. The experimental context was chosen to mirror the planned facial electromyography (fEMG) study as closely as possible. To shed light on the context specific interpretation of opponent smiles, we asked participants to explicitly rate each smile according to the following smile categories: appeasement (affiliation smiles), joy and schadenfreude. These categories were chosen according to the Simulation of Smiles Model (SIMS, Niedenthal et al., 2010), which distinguishes between smiles about rewards (joy), affiliation smiles, and dominance smiles, which perceivers recognize via the embodied simulation of the expression (i.e., facial mimicry), mind-reading of the expresser's state, and an evaluation of the current context. The social function of dominance smiles should in theory closely resemble the function of schadenfreude displays according to the schadenfreude as *"Social-Functional Dominance Regulator"* account (Lange & Boecker, 2019), which states that schadenfreude expressions function as a means to downregulate the social status of the target.

To investigate the perception of schadenfreude smiles within an aggressive competitive context, all smiles were integrated as target reactions into a modified competitive reaction time task (CRTT), which was designed as a successor of the Taylor Aggression Paradigm (TAP; Taylor, 1967). The CRTT allows for a measurement of aggression based on the choice of noise volume as an opponent punishment in a competitive task. The task procedure was used by Eder and colleagues (2021) to investigate the regulation of reactive aggression based on provocation and emotional target feedback. As a means of provocation we manipulated the opponent volume choices, with no-provocation conditions only featuring choices of low volumes (level 1 and 2), and provocation conditions featuring only high volume choices (level 4 and 5).

3.2.1 Method

Participants

Sixty participants volunteered for a payment of 10€. Each participant provided a written informed consent before participation. In line with our preregistered data exclusion rules (with the same exclusion criteria as for Study 1), ten participants were excluded because the provocation treatment was ineffective. The final sample consisted of 50 participants (13 male, $M_{age} = 24.45$, $SD_{age} = 7.61$). A sensitivity analysis indicated that this sample size had sufficient statistical power (1-beta = 0.80) for the detection of a small-to-medium effect ($d_z \geq 0.40$) in a two-tailed paired samples t-test. All participants provided informed written consent and all study procedures were in line with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Materials

Videos showing facial reactions of opponents in the competitive game were created in-house. Each video featured a white male with an age between 20 and 30 years wearing headphones seated in front of a grey wall (total amount of videos = 8, one per opponent). For the creation of a videotaped schadenfreude response, models were informed about the planned competitive setting with sound punishments of opponents and were instructed to smile maliciously while keeping direct eye contact with the camera. Videos were cut to 3 seconds, with each video starting with a neutral expression progressing into the schadenfreude response (for example stimuli see Fig. 6).

Figure 6

Example of the Opponent's Expression of Schadenfreude

Note. Screenshots from two video clips (duration: 3s), taken from the first 500ms segment at the left and the last 500 ms at the right.

The sound blast used for physical punishments consisted of a 3s recording of white noise taken from the Inquisit database (Millisecond Software, Seattle, WA, USA). The noise was calibrated to 75 decibels at the loudest setting (volume level 5) and the volume was decreased in steps of 5 dB for each volume level (1-5).

Procedure

The competitive reaction time task is framed as a reaction time game against another participant, who is only available over the internet. All participants were assured that while they can see their opponent, they themselves remain anonymous and cannot be seen by their opponent. Before starting a game round, each participant had to select a volume level for the noise their opponent will hear in case the participant wins, and that their co-player will do so as well. Punishment selection is carried out via a 5-point Likert scale, with a visual indication that 1 equals the lowest and 5 the highest volume. The selection was followed by a blank screen with a random duration between 1000 and 2000ms. After the wait period, participants were shown a red circle and a text cue ("ready?") for a random interval between 1000 and 2000ms, followed by the circle turning green and a text cue ("go!"). During the green circle phase, participants were tasked to click the mouse button, which was followed by the outcome display (3000ms). In a loss outcome trial, the participants were given loss feedback for 3000ms ("You have lost! Your opponent chose volume: X"), followed by a 3000ms noise blast. The noise blast was then followed by the opponent reaction video (either smile or neutral expression). After the opponent reaction, the participant was informed that the next trial would be starting or in case of the block being over, that the next opponent would be chosen. In win trials, the participant received a 3000ms win feedback ("You won! You have selected volume: X").

Each participant completed 12 CRTT blocks (4 provocation, 4 no provocation, 4 no provocation – random sequence) with 4 trials each. Within each block, the first trial was set to a random game outcome (win/loss), followed by 2 consecutive losses, followed by one win. The video showing the smiling opponent after a sound punishment of the participant was shown in the third trial; in the other losing trials the opponent reacted with a calm face to the punishment of the participant. Winning trials were not accompanied by opponent feedback. To mask the trial structure participants were also presented with random blocks without provocation (filler blocks that were not analyzed). During these filler blocks opponents always chose a low punishment volume and reacted with neutral facial displays. The order of block type and opponent was random. Each opponent was shown once.

Smiling videos were followed by three items that asked the participant for her interpretation of the opponent's smiling in respect to schadenfreude ("My opponent feels schadenfreude"), reward-based pleasure ("My opponent is enjoying that he won"), and appeasement ("My opponent makes an attempt to appease me"). The items were rated on a 7-point scale ranging from 1 ("I disagree a lot") to 7 ("I agree a lot"). An additional two items inquired after positive ("My opponent generally has a favorable attitude towards me") and negative attitudes of the opponent ("My opponent generally has a negative attitude towards me").

After each CRTT block, participants were asked to indicate their current mood, arousal and feelings of dominance on self-assessment manikin scales (SAM; Bradley & Lang, 1994). After completing 12 CRTT blocks, participants were debriefed, paid, and dismissed.

3.2.2 Results

Interpretation of the Opponent Smile

Figure 7 displays mean ratings averaged across smiling opponent and ratings for each individual opponent. Comparisons of mean ratings with t-tests showed that smiles of provoking opponents were rated as containing more schadenfreude ($M = 6.26$, $SD = 1.07$), $t(49) = 10.80$, $p < .001$, $d = 1.52$, and more enjoyment of winning the competition ($M = 5.91$, $SD = 1.32$), $t(49) = 3.41$, $p = .001$, $d = 0.48$, in comparison to smiles of the non-provoking opponent (M schadenfreude = 3.88, $SD = 1.54$; M joy = 5.21, $SD = 1.41$). Ratings of appeasement were generally on a low level and did not significantly differ (M provocation = 2.22, SD = 1.95; M no provocation = 2.60, $SD = 1.45$), $t(49) = 1.58$, $p = .120$, $d = 0.22$.

Opponent's attitudes were rated as more negative for provoking ($M = 5.33$, $SD = 1.29$, and $M = 2.20$, $SD = 1.02$) compared to non-provoking opponents ($M = 2.77$, $SD = 1.21$; and $M = 4.33$, $SD = 1.27$), $t(49) = 11.67$, $p < .001$, $d = 1.65$, and $t(49) = 10.37$, $p < .001$, $d = 1.47$.

Manipulation Checks of Provocation

Participants felt less pleasant during provocation blocks compared to non-provocation blocks, $t(49) = -5.07$, $p < .001$, *d* = 0.72. They were also more aroused after provocation, $t(49) = 7.75$, $p < .001$, $d = 1.10$, and reported less feelings of dominance, $t(49) = -4.39$, $p < .001$, $d = 0.62$. Please see Table 3 for descriptive values.

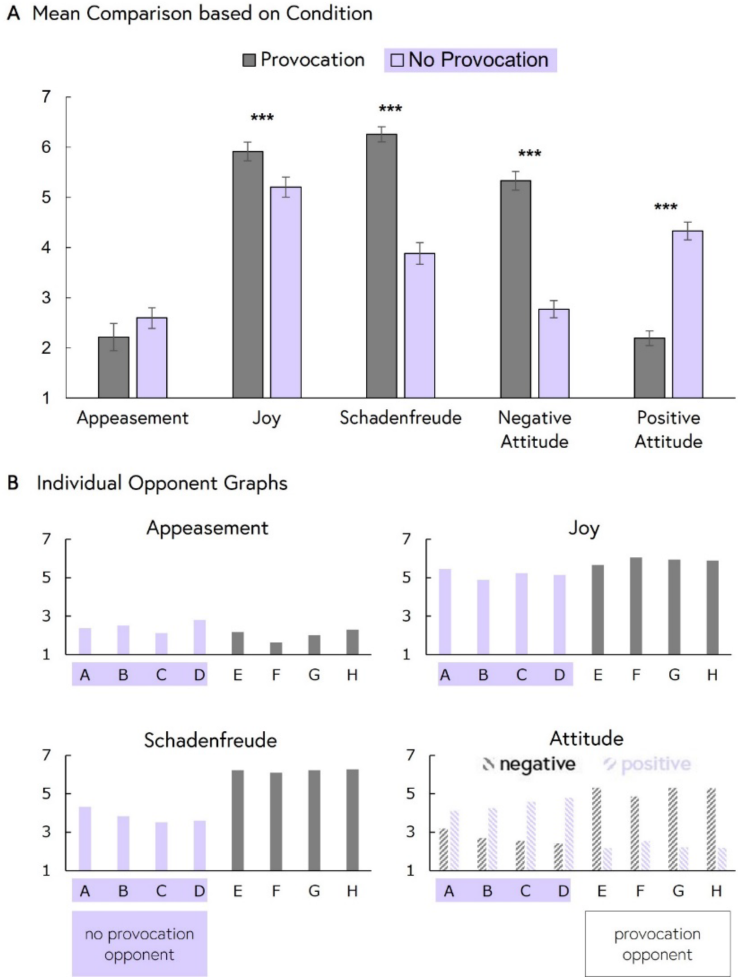
Participants were also angrier with provoking opponents, $t(49) = 11.93$, $p < .001$, $d = 1.69$. Provoking opponents were punished with higher noise levels than non-provoking opponents, $t(49) = 9.18$, $p < .001$, $d = 1.30$. In sum, these results indicate that participants were sufficiently provoked during provocation blocks.

Table 3

Punishment and Self-Reported Affect Means (*SD*s)

Condition	Punishment	Anger	Valence	Arousal	Dominance
provocation	3.03 (1.10)	3.57 (1.21)	3.49 (0.90)	2.78 (1.00)	3.20 (1.21)
no provocation	1.83 (0.86)	1.95 (0.82)	3.88 (0.74)	2.10 (0.80)	3.55 (1.15)

Figure 7



Note. Scales ranged from 1 (I do not agree) to 7 (I agree completely). Error bars represent standard errors of the mean (SEM).

3.2.3 Interim Discussion

In line with previous studies on the interpretation of smiles, data obtained in the pilot assessment indicates that the interpretation of opponent smiles is dependent on contextual social information. Although all smiling videos were created with the explicit instruction of a schadenfreude expression, involving direct staring and occasional head shaking as signs of disapproval, the ratings indicate that the attribution of intent behind these smiles was largely dependent on the provocation behavior of the opponent. While highly provoking opponents were assumed to be smiling because of schadenfreude, accompanied by a high rating of joy, the smiles of non-provoking opponents were attributed to joy most strongly. All of the opponent smiles were interpreted as low in appeasement. Ratings varied significantly dependent on context and less dependent on specific individual performances, as shown in the individual video ratings.

Context also varied the assumed attitude towards the participants in line with the provocation conditions. While highly provoking opponents were assumed to have a strong negative attitude towards their competition partners, the attitude of non-provoking opponents was assumed to be more positive. In summary, the pilot assessment indicates that the obtained videos of smiling reactions are interpreted as schadenfreude in the context of a competition paradigm and are therefore suitable as target reactions for the present EMG study, investigating the immediate facial reactions towards schadenfreude expressing opponents.

3.3 Study 1

The aim of our study was to provide an unobtrusive measurement of facial responses towards schadenfreude-displaying opponents in a competitive and anger-inducing social interaction. To manipulate schadenfreude responses during aggressive competitions, we used a modified version of a competitive reaction time task (CRTT) featuring video feedback by the (fictitious) opponent. Facial mimicry responses were measured with facial electromyography. We manipulated levels of provocation during the CRTT with pre-selected opponent responses punishing the participants with very intense versus very mild sound blasts after winning trials. We hypothesized that provocation will either weaken mimicry of smiling opponents, compared to mimicry of non-provoking opponents, or result in counter empathic facial responses, with smiling opponents inducing negative facial reactions. We defined counter-empathic responses as lower activations in zygomaticus major (ZM, mouth corner raising) muscle activity and higher corrugator supercilii (CS, brow furrowing) activation; in contrast, mimicry responses congruent with smiling were indexed by an increase in ZM activity and a decrease in CS activity.

3.3.1 Method

The experiment had a 2 (provocation: yes *vs* no) by 2 (emotional reaction: smiling *vs* neutral) within subjects factorial design. Materials, raw data, and analysis files can be accessed at https://osf.io/ykzvp/?view_only=da95e68f93ba4be18c17a5ea9c d9319f. The conditions analyzed in this paper were collected as part of a larger study on the affective responses toward target feedback reported by Mitschke & Eder (2021). The conditions, hypotheses and corresponding data analyses reported in the present chapter are original and were not included in the publication of Mitschke and Eder (2021).

Participants

Fifty-six participants were recruited from the online participant management pool of the University of Würzburg. Each participant received 15€ as compensation for participation. After data inspections and exclusion, the final sample consisted of 39 participants ($M_{age} = 23.64$, $SD_{age} = 3.70$, 5 male). Eleven participants had to be excluded due to technical difficulties (e.g., electrode loss), excessive movement or unrelated facial responses such as yawning. An additional six participants were excluded due to not being sufficiently provoked, indexed by a comparison of mean punishment choices averaged for each condition. For a similar exclusion procedure see also Eder et al. (2020) and Mitschke et al. (2021).

All participants provided informed written consent and the study procedures were approved by the local ethics committee of the university (GZEK 2020- 74). A sensitivity analysis performed with G*Power (Version 3.1.9.2) indicated that the sample size had sufficient statistical power (1-beta = .80) for the detection of an ANOVA effect of $f \geq .24$, with the correlation among repeated measures set to 0.2 and the nonsphericity correction to 1.

Stimulus Materials

Videos materials showing opponent smiles and noise blasts were identical to those used in the aforementioned pilot study. We added a measurement of trait empathy, which was assessed with the German version of the Interpersonal Reactivity Index (IRI; Davis, 1983), called SPF (Paulus, 2009), for exploratory reasons.

Procedure

Muscle activity was measured using bipolar Ag/AgCl surface electrodes (4 mm, distance 1.5cm) attached to the left side of the face and behind the right ear for a reference signal. Electrode placement was carried out in line with the guidelines for human electromyographic research (Fridlund & Cacioppo, 1986). Specifically, electrodes were placed to measure activity of the ZM that is activated while raising the mouth corners during smiling, and the CS that is activated during the lowering of the brows during frowning. Muscle activations were recorded with a 16-channel amplifier (V- Amp, Brain Products, Gilching Germany).

After placement of the electrodes, participants were seated in a soundproof chamber and informed that they would play an online competitive game against an opponent situated at a different university. Task instructions highlighted that they were randomly assigned to a condition in which they could see a video stream of the (fictitious) opponent during the competition, whereas the opponent could not see them. Participants were debriefed about the deception and the computer-generated fictitious opponent at the end of the experiment.

The CRTT consisted of 48 game rounds (blocks). Each block consisted of 5 game trials against the same opponent. Participants played against 12 different opponents, which means that they played four game blocks against each opponent in a single session. The opponent for a game block was selected randomly but could not repeat in two consecutive blocks. Thirty-two game blocks were 'no provocation blocks' with exclusively low volume punishments (level 1 and 2), sixteen game blocks were 'high provocation blocks' that always featured high volume punishments (volume levels 4 and 5). The no provocation blocks were split into random filler and measurement blocks. Measurement blocks always consisted of the same win/loss order (3 losses, then 2 wins). Random filler blocks consisted of random distributions of wins and losses and only featured neutral expressions. These blocks were included to conceal the fixed win/loss ratio in measurement blocks and were not analyzed.

Win/ loss outcomes were shown directly after the reaction task. If no response was collected, the procedure was repeated. When the participant lost the CRTT, they heard the sound blast selected by the opponent. The sound blast was followed by a 3 second blank screen to allow potential muscle reactions due to the noise blast to subside and provide a baseline period. After the 3s blank screen, a video showed the opponent who reacted to the sound punishment of the participant either with smiling or with a calm (neutral) face. When the participant won the game, the opponent reacted with expressions of pain or with neutral expressions. Analyses of participants' facial activities during viewing the opponent's reactions in win trials are reported in Mitschke and Eder (2021). The present article analyses participants' facial muscle activity only during the observation of opponents within loss trials when they observed a smiling opponent.

After each block, participants rated how angry they were about the opponent (5 point scale ranging from 1 'not at all' to 5 'very') and their current affective feelings (pleasantness, arousal, dominance) on SAM scales (Bradley & Lang, 1994). After the CRTT, participants completed the SPF questionnaire. Then, they were debriefed, paid, and dismissed.

Design and Analyses

Raw fEMG signals were filtered (20 Hz low cutoff filter, 499 Hz high cutoff filter, 50 Hz notch), full wave rectified, segmented (−200 ms prior to video onset to 3,500 ms post onset), baseline corrected (via subtraction of facial activity during the 200 ms period prior to video onset), averaged per condition, and cut into 500 ms time bins (resulting in seven time bins.)

The processed fEMG data were analyzed with a linear mixed model (LMM) in JAMOVI (using the GAMLj module, version 1.0.7; The Jamovi Project, 2021). The model included

Provocation (high vs low), *Emotion* (neutral vs smile), *Time* (500 ms intervals across 3500 ms) and interactions thereof as fixed effects. Subject was included as random effect. As dependent variables the baseline-corrected EMG responses for each muscle (ZM, CS) were analyzed separately, resulting in one model per facial muscle. Significance was calculated using the Satterthwaite's method. The model specification was:

MuscleResponse ~ 1 + Provocation + Emotion + Time + Provocation:Emotion + Provocation:Time + Emotion:Time + Provocation:Emotion:Time+(1 | Subject)

3.3.2 Results

fEMG Data

Figure 8 shows the time based average of participants' facial muscle activity during the observation of a smiling versus calm opponent divided by provocation levels (high vs low).

Zygomaticus Major (ZM)

In the omnibus test, the fixed effect of *Emotion*, $F(1,1026) = 77.83$, $p < .001$, *Time*, $F(1,1026) = 8.04$, $p < .001$, and the *Emotion x Time* interaction, $F(1,1026) = 4.88$, $p < .001$, were significant. ZM muscle activity was significantly larger when seeing the opponent smiling compared to calm, ($b = 49.85$, $SE = 5.65$), $t(1026) = 8.82$, $p < .001$. The magnitude of the ZM increase was not modulated by *Provocation*, $F(1,1026) = 1.45$, $p = .229$, *Provocation* x *Emotion* interaction, $F(1,1026) = 0.89$, $p = .346$, or *Provocation* x *Emotion x Time* interaction, $F(1,1026) = 0.219$, $p = .971$.

Follow-up post hoc comparisons of *Time* across both provocation conditions, revealed a significant increase in ZM activity during time bins 3-7 (1000 to 3500ms, all ts ≥ 2.04, all ps $\leq .040$). This timing roughly corresponded with the mean timing of the opponent's peak smiling expression in the video clips (see Figure 8 for a time progression graph).

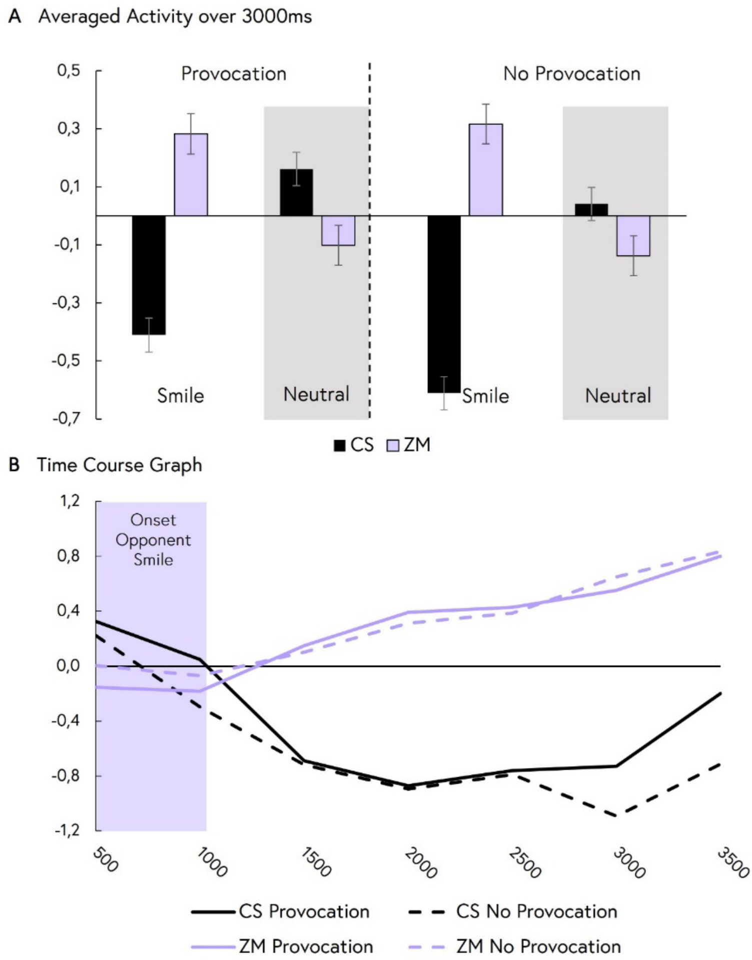
Corrugator Supercilii (CS)

The omnibus test yielded significant main effects of *Emotion*, $F(1,1026) = 190.392$, $p < .001$, *Time*, $F(1,1026) = 18.03$, $p < .001$, *Provocation*, $F(1,1026) = 15.09$, $p < .001$, and a significant *Emotion* x *Time* interaction, $F(1,1026) = 9.92$, $p < .001$. CS muscle activation significantly decreased when seeing a smiling opponent compared to a calm opponent, ($b = -77.14$, $SE = 5.59$), $t(1026) = 13.80$, $p < .001$. Interactions between *Provocation* and *Emotion*, $F(1,1026) = 0.16$, $p = .694$, as well as *Provocation* and *Time*, $F(6,1026) = 0.56$, $p = .770$, and the three-way interaction between *Provcation, Emotion* and *Time*, $F(6,1026) = 0.58$, $p = .750$, all yielded non-significant effects.

Follow-up comparisons within the smiling condition revealed that the CS muscle decrease was significantly larger when seeing non-provoking opponents smile compared to provoking opponents ($b = -23.92$, $SE = 7.90$), $t(1026) = -3.03$, $p = .003$.

Analyses of simple effects of *Time* within the emotion condition smiling, separately for each provocation condition, revealed significant decreases in CS activity during time bins 2-7 (from 500 to 3500ms) in both provocation conditions respectively (all bs $\geq$ -79.47, all ps $\leq$ .001).

Figure 8



Note. Facial muscle activations (z-transformed) as a function of the opponent's facial display (smiling, calm) following sound punishments of the participant and provocation levels.

Punishment Levels

Difference scores were calculated for each provocation condition via subtraction of volume choices before the opponent's smiling response from volume choices in the consecutive trial. Analogous difference scores were calculated for blocks with neutral opponent reactions (i.e., before and after seeing a calm opponent reaction to sound-blasting the participant). Positive values indicate an increase in punishment following the opponent reaction, whereas negative scores indicate a decrease. A 2 (provocation) by 2 (emotion) rmANOVA, revealed a significant effect of *Provocation*, $F(1,38) = 5.84$, $p = .021$, $\eta_p^2 = .133$ and no significant main effect of *Emotion* $F(1,38) = 2.78$, $p = .104$, $\eta_p^2 = .068$, or interaction between *Provocation X Emotion*, $F(1,38) = 2.83$, $p = .101$, $\eta_p^2 = .069$.

Follow up t-test comparisons within each provocation condition resulted in significant differences following smiling and calm opponents. In provocation blocks, participants opted to punish smiling opponents more in the next trial ($M = 0.14$, $SD = 0.36$) compared to non-smiling opponents ($M = -0.20$, $SD = 0.59$), $t(38) = 2.85$, $p = .007$, $d_z = 0.59$. In no provocation blocks, sound blasts selected for administration to calm ($M = 0.26$, $SD = 0.98$) and smiling opponents ($M = 0.24$, $SD = 0.45$) did not differ significantly, $t(38) = 0.15$, $p = .880$.

Manipulation Check of Provocation

Punishment was more intense after provocation, indexed by higher volume choices in punishment blocks compared to non-provocation, $t(38) = 8.13$, $p < .001$, $d = 1.30$ and significantly more feelings of anger towards the opponent following provocation blocks, $t(38) = 9.18$, $p < .001$, $d = 1.47$. In addition, participants felt more unpleasant, $t(38) = 4.90$, $p < .001$, $d_z = 0.79$; more aroused, $t(38) = 4.34$, $p < .001$, $d_z = 0.70$; and less dominant, $t(38) = 3.63$, $p < .001$, $d_z = 0.58$, following provocation. See Table 4 for a summary of the descriptive data of these measurements.

Table 4

Means (SD) manipulation check of provocation

	Noise Level	Valence	Arousal	Dominance	Anger
provocation	2.99 (1.14)	3.63 (0.72)	2.12 (0.81)	3.49 (1.10)	3.08 (1.05)
no provocation	2.00 (0.83)	3.81 (0.70)	1.79 (0.66)	3.71 (1.10)	1.93 (0.69)

Note. Ratings are based on 5-point scales with 1 indicating no/the least amount and 5 indicating the largest amount, except for valence which ranged from 1 negative to 5 positive.

3.4 Discussion

We investigated spontaneous facial reactions and subsequent punishment choices towards smiling opponents and the interpretation of contextualized smiles during a competitive aggressive interaction. Seeing the opponent smile after one's own punishment intensified punishments selected for the next trial, showcasing the escalation potential of schadenfreude for aggressive interactions. Contrary to the hypothesis of a counter-empathic facial response, however, facial EMG recordings demonstrated smile mimicry. Results clearly indicated an increase in ZM muscle activation, while CS muscle activity decreased, a pattern that is indicative of smiling. Onsets of the individual muscle reactions were coordinated, roughly appearing at the same time and corresponded to the timing of the peak of the opponent smiling reaction. Concerning the attenuation hypothesis, we observed a partial attenuation of smile mimicry towards highly provoking opponents, which was only present in the CS muscle relaxation and not in the ZM muscle activation. These results closely resemble the muscle patterns observed after competition priming obtained by Weyers et al. (2009). Smile mimicry was affected by provocation, evident in the significant differences in CS decrease. Decreases in CS muscle activity have been linked to increased positive affect (J. T. Larsen et al., 2003). The attenuation of CS relaxation could hence be interpreted as a less inherently positive response towards highly provoking opponents.

The observation of smile mimicry during aggressive competition in absence of any counter empathic muscle reaction stands in contrast to previous studies which have observed a full suppression of mimicry or counter empathic reactions towards disliked others and in competitive settings (Blocker & McIntosh, 2016; Lanzetta & Englis, 1989; Likowski et al., 2008). Apart from the differences in results, these studies differ in methodological aspects from our study. We chose to embed our measurement into an interactive game set up with dynamic video displays of smiles. Comparisons of mimicry during static and dynamic emotion displays indicate that dynamic facial displays evoke stronger mimicry responses (Sato et al., 2008). However, Blocker et al. (2016) used dynamic naturalistic displays of human models and observed an absence of smile mimicking behavior for smiling persons previously paired with negative behavioral descriptions, presumably evoking negative attitudes towards the target. Hence, the usage of dynamic expressions as reactions is not fully sufficient to explain the differences in the obtained results from previous studies.

Another profound difference between the aforementioned studies and the paradigm of the present study lies in the task context in which the facial displays were observed. All above mentioned studies used a passive viewing task in which no interactive context was utilized between the observer and the emotion expressing person or avatar. The absence of interactive context could explain the previously observed absence of mimicry. Mimicry responses are goal and task dependent (Murata et al., 2016) and therefore the engagement in the interaction may change the modulation of mimicry according to the perceived need for understanding the interaction partner (Seibt et al., 2015).

Monitoring and understanding the opponent may give the perceiver advantages when it comes to interactive competitions. The simulation of emotional states via mimicry have been shown to aid understanding, effectively allowing the executor to recognize emotions faster (Wood et al., 2016). In the case of the CRTT setting the observer likely anticipates future interactions with the opponent, as each opponent is encountered multiple times. Additionally, decision making during the CRTT, has been shown to involve mentalizing networks, indicating a cognitive process to estimate not only personal goals and states but also predicting the potential outcomes for the opponent (Chester & DeWall, 2016). The need for predicting the opponents state could potentially favor the occurrence of rapid mimicry to aid in emotion recognition to ease the estimation of the opponent state.

The quick onset and comparatively low attenuation of the mimicking response (only affecting one muscle site) suggests a fact acting imitation process. Smiling, although in our case with a bad intention towards the observer, may still be processed as a strong affiliative signal (Knutson, 1996), at least before evaluating the personal negative meaning for the observer. Since the present study analyzes a relatively short time frame (3500ms) directly during the observation of the smiling response, it remains possible that negative reactions towards positive displays, classified as counter-empathic responses, occur after the measured time frame and are therefore not captured within our analysis and experimental structure.

We did not observe immediate reversals in facial reactions towards smiling opponents, and only observed negative facial reactions towards highly provoking opponents displaying a neutral facial expression. Future studies investigating the occurrence of counter empathy may profit from longer time frames and a detailed differentiation of immediate and later progressing muscle onsets to differentiate between facial mimicry and affective reactions towards the target. Strong affiliative signals, such as smiles, may result in initial fast mimicking responses due to the observation and repeated execution of smile imitations over the lifespan (Heyes, 2001). Especially reactions toward smiles of disliked others could serve as an interesting venture point for future studies to investigate the underlying inhibition processes modulating automatic mimicry responses based on social aspects.

The mimicking of antagonistic facial displays has also been indicated to deteriorate social quality of interactions (Hess, 2021), hence a mimicking of schadenfreude could also serve an aggressive communicative function in reacting in a congruent antagonistic manner. Observers could have mirrored the malicious smile to signal their own aggressive intentions or save face during the encounter. If schadenfreude functions as a dominance regulator, participants may responded with equal dominance, as being nonresponsive towards dominance displays could be viewed as a signal of submission (Mazur & Booth, 1998). Further studies are needed to differentiate between these alternative functional explanations of opponent smile mimicry.

Apart from dominance regulation, opponent smiles could also have been interpretated as signaling appeasement. Evaluating the opponent smiles as appeasing could have led the participants to communicate congruent smiles as a form of peace making. However, behavioral data indicated that smile responses were followed by an increase in punishment, which would be detrimental to de-escalation. To rule out smile interpretations deviating from schadenfreude, we conducted the pilot study. Results from the explicit rating indicate that participants are prone to attribute smiles from highly provoking opponents as schadenfreude and enjoyment but not as an action that is attributed to reaffiliate via appeasement.

The present study has some important limitations. First, data about the attribution of smiles in the pilot assessment was obtained from an unrelated sample of participants via explicit ratings. We can therefore only make assumptions about the potential to perceive opponent smiles as schadenfreude in a competitive setting but cannot provide direct evidence that the participant sample in Study 1 categorized smiles in the same manner. Nevertheless, multiple studies have shown that context reliably affects the attribution of smiles (Hess & Hareli, 2017; Rychlowska et al., 2019), thus it is unlikely that participants misinterpreted the opponent smiles, given the competitive context as well as the reported pilot study.

Second, most participants were females, and we cannot generalize our findings to males. As only 5 male participants completed the fEMG study, it is not feasible to run gender differentiating analysis within the current sample of participants. Investigation of the neural circuitry of the execution of smile mimicry and recognition of smiles suggest differences in the processing of mimicry based on perceiver gender (Korb et al., 2015). Females have been shown to possess higher implicit affiliation motivation (Drescher & Schultheiss, 2016), which could theoretically result in a higher potential for mimicking responses.

In summary, we provided evidence that previous provocation combined with opponent smiling fosters the attribution of smiles as schadenfreude and that this hostile opponent condition is accompanied by an increase in aggressive motivations, potentially escalating aggressive interactions. Our fEMG results indicate that schadenfreude smiles were mimicked, and opponent mimicry were partly attenuated due to prior provocation.

The coordinated smiling with the opponent did not hinder further aggressive responding, which favors the assumption that mimicking behavior during the aggressive competition context was not used to communicate affiliation towards the opponent but may serve a simulative function to allow for a better recognition and subsequent attribution of opponent smiles or serves as a gesture of dominance.

Irrespective of prior provocation, perceivers mimicked the opponent smiles in a fast and time congruent manner, indicating that counter empathic responses did not occur immediately and may require processes that are based on later appraisals or subsequent corrective actions, reversing the initial smiling response. The presented results potentially provide an interesting avenue for future studies to investigate the different potential functions of smile mimicry in negative social contexts and their relation to counter-empathic responding.

CHAPTER IV

Chapter 4 – Remembering the Enemy

Empirical Synopsis

Social perceptions are anything but static. The processing of new social information leads to dynamic updates of impressions and representations of others. Two experiments (N = 592) investigated how aggressive competitive interactions affect the visual representation of opponents using a reverse correlation task. In a first experiment phase, participants interacted with an aggressively behaving opponent and were subsequently asked to reconstruct the opponent's face from memory. An independent participant sample naive to the foregoing interactions then rated the generated images of the opponent on several trait dimensions. Reconstructed images of revenge-provoking opponents were rated as less trustworthy and likable. A second study manipulated emotional reactions of the opponent to (retributive) punishments. Reconstructed faces of opponents who expressed suffering after revenge taking were rated by an independent sample as more trustworthy, less hostile, and less dominant compared to opponents who reacted with anger to punishments. These findings demonstrate that social processing influences how an aggressive interaction partner is perceived and visually remembered, and that both, aggressive behaviors, and emotional reactions during the social interaction become integrated into the visual representation of the aggressor.

4.1 Introduction

Social interactions and their outcomes affect how interaction partners are perceived and remembered. When a person is mistreated by another person, she likely analyzes the social behavior of the other person, making inferences about potential malevolent intentions and antisocial traits that can explain the aggressive behavior (Hareli & Hess, 2010; Trope, 1986). These social inferences and moralistic judgments color how the aggressor is evaluated and remembered by the target of the aggression, which in turn influences future interactions with that person. The present study investigated whether these interactive experiences also affect the target's visual representation of the perpetrator.

Only few studies investigated how social interactions modulate the visual representations of interaction partners from memory. Mental representations of persons integrate various information sources, such as perceptual features, traits, evaluations, and social categorizations (Wyer, 2007). Impressions of interaction partners are subject to continuous updating whenever new information is available (Mende-Siedlecki, 2018). Supporting the existence of constant dynamic update processes during social interactions, Chang and colleagues (Chang et al., 2010) demonstrated that the experience of behavior of an interaction partner leads to trial by trial updates of subsequent reciprocal trust decisions. The authors investigated the processes behind trustworthiness decisions and judgements in an interactive paradigm, showcasing that trust judgements are constantly updated, resulting in dynamic changes in trust behavior following each single interaction.

Learning about the behavior of a person not only updates impressions but has also been shown to selectively affect visual perception and memory for the faces of interaction partners (Davis et al., 2010; Mattarozzi et al., 2019). Falvello and colleagues (2015) showed that when pairing faces with written statements of positive or negative behaviors, persons previously paired with a single positive behavioral statement were rated as more trustworthy than those previously combined with negative behavior. The combination of faces with behavioral information also boosts face memory (Mattarozzi et al., 2019), with negative behaviors combined with untrustworthy faces producing the strongest memory advantages (Rule et al., 2012).

On top of behavioral information, the emotional state of the perceiver can bias the processing of information towards a congruency with the person's emotional state. During aggressive interactions, targets of aggression likely experience negative affect due to prior provocation or frustration which, in line with a negativity bias, directs their attention towards the negative aspects of the ongoing interaction and their interaction partner.

Biases in the perception and encoding of facial expressions may play a pivotal role in the maintenance of aggressive behavior. Research investigating perceptual biases during aggressive acts indicates that aggressive individuals show perceptual biases towards aggressive attributions of intents and emotional states. This so-called hostile-attribution bias (Nasby et al., 1980) results in the attribution of negative intents and has been shown to bias the categorization of ambiguous facial expressions towards anger (Mellentin et al., 2015). These studies indicate a potential influence of chronic aggressive motivations on the encoding and evaluation of faces, which may be part of a self-sustained circle between aggressive expectations and responses, ultimately escalating aggressive interactions.

Balas and Thomas (2015) demonstrated that competitive social contexts bias the reconstructive memory for faces towards higher facial width to height ratios (fWHR), which are generally associated with higher judgments of dominance or aggression. The authors staged competitive and cooperative interactions with confederates, and later provided participants with the task to assemble facial portraits using composite parts of their interaction partners' face. Participants were instructed to arrange the eyes, mouth, and nose of their interaction partner from memory into a face. Composite arrangements were then measured and judged by an unrelated sample of participants. The authors demonstrate that social perception is not only reliant on strict feed-forward processing but also affected by higher level top-down estimates that uniquely shape the perception and memory of others. In the case of competition, the higher-level social context biases the representation towards facial configurations that are associated with conflict, resulting in changes that are congruent with social judgements of aggression and competitiveness.

A different method to visualize mental representations of faces was introduced by Dotsch and colleagues (2008), the so-called reverse correlation method. This approach uses the choice between multiple images overlaid with random pixel noise to approximate the "mental" image of a person reconstructed from memory. Using this method, studies demonstrated that prejudiced individuals reconstructed faces of out-group members that looked more criminal and less trustworthy to another, non-prejudiced sample of participants. This research indicates that negative attitudes can be transferred onto mental visual representations. The bias in the representation of out-groups has also been demonstrated with minimal group designs, suggesting that social context, such as membership in a group, can affect subsequent visual representations of persons (Ratner et al., 2014).

The present research

The present research investigated whether aggressive social interactions systematically affect the visual representation of the opponent's face. The visual image of the aggressively behaving opponent, as reconstructed by the reverse correlation method, was hypothesized to appear more unfavorable to an independent participant sample compared to the image of a non-aggressive opponent, as indexed by increased image ratings of untrustworthiness, hostility, and dominance. For the manipulation of the opponents' aggression levels, a competitive reaction time task (CRTT) was used in which the fictitious opponent physically punished the participant with allegedly unwarranted selections of extremely severe or very mild noise blasts (resulting in high vs low provocation levels, respectively).

To quantify the effect of provocation on the visual representation, an independent sample of participants, naïve to the aggressive interaction contexts, rated the faces reconstructed from the image selections of the first sample. Two social dimensions are particularly influential for social face perception and were of primary research interest: judgments of trustworthiness and dominance (Oosterhof & Todorov, 2008). Trustworthiness is linked to social perceptions of warmth, friendliness, and predictions of prosocial intentions of the other person (van ’t Wout & Sanfey, 2008). Perceptions of dominance, or the competency of another person, are used for inferences about a person’s capacity for benevolent or malevolent actions (Sutherland et al., 2016). Additionally, the punishment behavior of the aggressive opponent should be associated with increased levels of perceived aggression, which is visually often correlated with male facial characteristics, such as the width to height ratio and brow protrusion (Balas & Thomas, 2015; Carré et al., 2009; Geniole et al., 2012; Lefevre & Lewis, 2014).

4.2 Study 1

Participants in the CRTT played either against an aggressively behaving or a non-aggressive opponent (random assignment). The aggressive behavior of the opponent was expected to lower the estimated trustworthiness and likability of the interaction partner. The repeated successful use of aversive punishment should also increase the estimate of dominance, as opponents are experienced to be highly capable of repeatedly carrying out the punishment. Based on the aforementioned predictions, we hypothesized that faces reconstructed from the provoking opponent should be perceived as more dominant and masculine, and less trustworthy and likable, compared to the non-provoking opponent.

4.2.1. Method

We report how we determined our sample size, all data exclusions, all manipulations, and all pre-registered measures in the study. Materials, data sets, the programming code, and analysis documents can be accessed at https://osf.io/7nup9/?view_only=2bd6befa4e8b46caaa6296b7aa95e9c9.

The study had two phases: (1) an image creation phase in which participants interacted with a provoking or non-provoking opponent (between subject) and created classification images (CI) of the opponent. (2) A rating phase in which a separate group of participants rated the reconstructed face images (CI) in reference to likability, trustworthiness, dominance, and masculinity.

Participants

The image creation sample (Sample 1) consisted of 83 participants (45 female, M_{age} = 26.88, *SD* = 8.13). Participants were recruited via the subject management pool of the psychology department and were paid 9€ for participation. Participants were randomly assigned to one of two conditions (provocation: *n* = 42; no provocation: *n* = 41). Previous studies using the reverse correlation method demonstrated that a sample size with *n* = 20 participants per condition is sufficient for the detection of visual changes from reconstructed images (Dotsch & Todorov, 2012; Oliveira et al., 2019). We preregistered a minimum of 40 participants per condition given the uncertain effectiveness of our aggressiveness manipulation. A sensitivity analysis showed that the sample N = 83 was sufficiently powered (1-beta = 0.8) to detect medium-sized effects ($d \geq 0.62$) in a two-tailed independent t-test with alpha = 0.05.

The image rating sample (Sample 2) consisted of 134 participants (73 female, M_{age} = 29.02, *SD* = 10.56), who were recruited via the internet platform Prolific. Each participant was paid 1£. For the rating procedure, we preregistered a minimum sample of 70 that had sufficient statistical power (1-beta = 0.80) for the detection of small-to-medium effects $d_z \geq 0.32$ in a dependent sample t-test with alpha = .05 (two-sided). Due to a grossly unbalanced gender ratio after the recruitment of n = 70, we collected additional data for a balanced gender ratio (in line with our preregistration). A sensitivity analysis with the final sample size

N = 134 revealed that this sample size resulted in sufficient power for the detection of small effects $d_z \geq 0.24$.

Materials

The sound blast in the CRTT was a 3s white noise taken from the Inquisit Database (Millisecond Software, Seattle, WA, USA). The loudness of the sound blast could range from 55 dB (level 1) to 75 dB (level 5) subdivided in 5 dB steps.

The image depicting the opponent was taken from the Radboud Faces Database and showed a male Caucasian with a neutral expression (see Panel A in Figure 9). The image had a mean attractiveness rating of $M = 2.5$ (medium) and a valence rating of $M = 3.2$ (neutral) in normative image ratings on 5-point scales (Langner et al., 2010). The base image was cropped and resized to 512 x 512 pixels, converted to greyscale, and filtered using a Gaussian blur function. A total of 1,280 noise images were created by overlaying a random noise pattern and inverted noise counterparts on the base image using the R package rcicr (Dotsch, 2016). Panel B in Figure 9 shows examples of the noise overlaid pictures.

Following the image classification task, the noise images selected by the participant were averaged and the averaged noise pattern was laid over the original base image, resulting in a CI for each condition (provocation *vs* no provocation). The same procedure was applied to non-selected noise images for the construction of so-called *Anti-CIs*. The (Anti-)CI construction was done via the R standard procedure implemented in the rcicr package (version 0.3.4.1.) using default settings (Dotsch, 2016; for a detailed guideline see Brinkman et al., 2017).

Procedure

Before the CRTT, participants rated their current feeling state on the dimensions pleasantness, arousal, and dominance using Self-Assessment Manikin scales (SAM; (Bradley & Lang, 1994).

CRTT. A (Taylor, 1967) aggression paradigm was used to create an aggressive interaction context. Participants played 25 trials against a (fictitious) opponent who was introduced to the participant with a display of the base image at the start of the game. CRTT instructions were to click the mouse button as quickly as possible when a (preparatory) red circle turned into green (go signal). The player who (allegedly) clicked the mouse last would hear a noise blast of the intensity selected beforehand by the other player. However, the outcome of the competitive task was actually rigged: Participants won the competition in half of the CRTT trials—allegedly sound blasting the fictitious opponent—and they lost the competition in the other half of the trials, resulting in the sound blasting of the participant with a noise blast supposedly selected by the opponent. The sequence of winning and losing CRTT trials was random.

In each CRTT trial, participants first selected the volume of the noise blast that should be delivered to the opponent if the participant wins the RT game. After the volume selection, the opponent image with the statement "Wait for Opponent Input" were shown for a random duration between 1,500-2,000 ms. In the provocation condition, the (computer-generated) opponent consistently selected high volumes for the noise blasting (levels 3-5). Previous research showed that this treatment provokes retaliatory aggression and unpleasant feelings (anger) in the participant (Eder et al., 2020; Mitschke & Eder, 2021). In the no provocation condition, by contrast, the opponent consistently selected low volumes (levels 1-2).

After completing 25 trials of the CRTT, participants were asked to rate their current feelings on the SAM scales, feelings of anger ("How angry are you with your opponent?") and likability of the opponent ("How likable is your opponent?") on 5-point scales (1 = not at all, 5 = extremely).

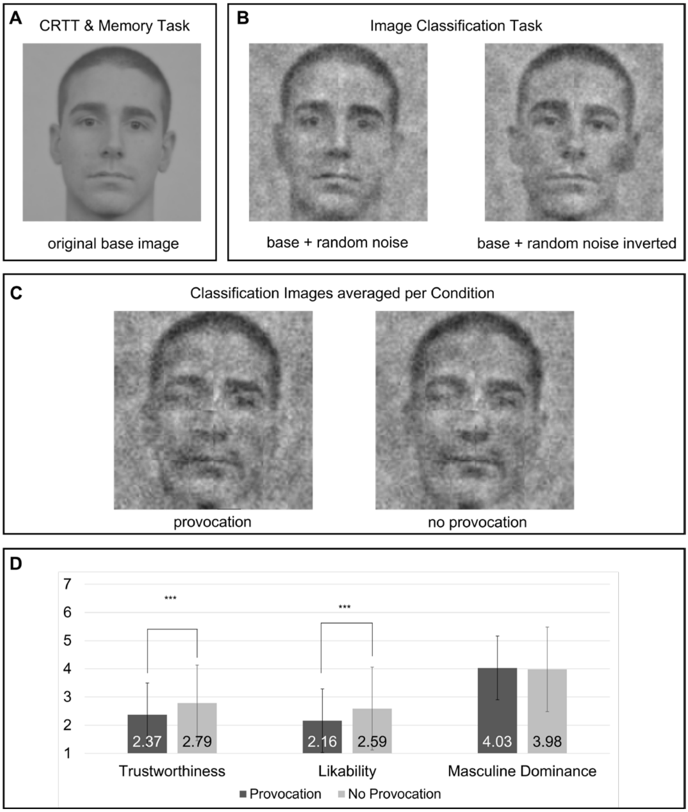
Memory Phase. After the rating, the picture of the opponent was shown for two minutes. During this time period, participants were to memorize the face as best as they can ("Please memorize your opponent's face. We will ask you further questions about your opponent later. You have two minutes.“). The memory phase was followed by a neutral filler task that consisted of unrelated image ratings of lay person drawings (without depictions of faces or persons). The filler task took approximately 15 minutes and, following the procedure of previous studies using the reverse correlation method (Karremans et al., 2011), it was included to extend the time between memorization and memory recall phases.

Image classification phase. After the filler task, participants completed the reverse-correlation image classification task (Brinkman et al., 2017). This task was designed to capture participants' visual memory of the opponent's face. Each trial showed two noise images (noise image on the left and inverted noise image on the right, with the sequence of image pairs randomized), and participants were instructed to select the image that best resembles the opponent previously seen (two image forced choice procedure, 2IFC). An image was selected with pressing the left or right arrow key corresponding with the image position. After completion of the 2IFC, the participant was debriefed, thanked, and dismissed.

Image rating phase. The CIs produced by Sample 1 were averaged per condition and rated by Sample 2 that was naïve to the previous study phases. The rating study was done online using the platform Prolific. For a familiarization with the pictorial material, participants first saw all 2 CIs and 2 Anti-CIs in a grid layout next to each other. This was done to ensure that participants are able to notice similarities and subtle differences between the images (for a similar procedure see Dotsch et al., 2008). The grid was shown for 1 minute, followed by single presentations of the CIs and Anti-CIs in random order. Participants rated each CI on dominance, masculinity, trustworthiness, and likability on a scale ranging from 1= *very little* to 7 = *very much* (*"Please indicate how* [dominant/masculine/trustworthy/likable] *this picture looks"*). After the rating task, the CIs of both conditions were directly compared on each dimension using the 2IFC ("Please indicate which of these two faces looks more … [dominant/masculine/trustworthy/likable]"). 2IFC comparisons and ratings of the random noise image and Anti-CI's were collected for exploratory analyses reported in Appendix B.

Figure 9

Materials, Classification Images, and Ratings for each Condition in Study 1



Note. Panel A shows the base image for the generation of noise images. This image was shown at the start of the CRTT and in the memory phase. Panel B depicts examples images consisting of the base image overlaid with noise and the inversion of the noise pattern. Panel C depicts the resulting classification images that were created by averaging noise image choices from Sample 1 for each condition separately. Panel D shows mean CI ratings obtained from Sample 2 (error bars denote *SD*s, ***$p < .001$).

4.2.2. Results

Before analyzing each item separately, a correlation analysis between all items was carried out. Items that were highly correlated (above r = .7) were added into an index via a mean score built from both items. The respective mean score index was then analyzed. See Appendix B for a correlation matrix including all measurements and the original analysis without index scoring.

An index was computed for the items assessing dominance and masculinity (item correlation $r(133) = .75$, $p < .001$), we will refer to this dimension as the masculine dominance index from now on. Three dependent t-tests were carried out to compare each judgment dimension (trustworthiness, likability, masculine dominance index) between images created based on provoking and un-provoking opponents and tested against a Bonferroni-adjusted alpha value of .016 (0.5/3).

Image Ratings (Sample 2). The CIs for each condition differed significantly in ratings of trustworthiness and likability (see Figure 9 Panel D). The CI created by the provoked group was rated as significantly less trustworthy, $t(132) = -3.92$, $p < .001$, $dz = 0.34$, and less likable, $t(132) = - 4.03$, $p < .001$, $dz = 0.34$, in comparison to the CI created by the unprovoked group. A t-test comparison of the masculine dominance index between both CI images revealed no significant differences, $t(132) = 0.47$, $p = .642$, $dz = 0.04$.

As an exploratory additional analysis for Study 1, we asked the rating sample to directly compare both classification images (CI) obtained from each condition. CI images were shown next to each other, and participants were asked which of the two images was more trustworthy (likable, dominant, masculine; 134 choices total). Results obtained from the direct comparison mirror those obtained in the scale-based ratings. Table 5 presents the results from this direct comparison task. Exploratory analysis comparing both Anti-CIs can be found in the Appendix B (Table S1).

Manipulation Checks of Provocation (Sample 1). Ratings of anger feelings and likability of the opponent after the CRTT were analyzed as a function of provocation. Comparisons with independent samples t-tests revealed that the provoking opponent was liked less than the non-provoking one, $t(81) = -4.38$, $p < .001$, $d = -0.96$. Participants assigned to the condition with the provoking opponent also expressed significantly higher levels of anger, $t(81) = 6.94$, $p < .001$, $d = 1.49$, and selected more severe noise blasts for sound punishment, $t(81) = 3.66$, $p < .001$, $d = 0.80$. Descriptive statistics are presented in Table 6.

Table 5
Binary choice comparisons via binomial tests (Sample 2).

	CI Type	Count	Total	Proportion	p
trustworthiness	Provocation	38	134	0.284	< .001
	No Prov.	96	134	0.716	< .001
likability	Provocation	29	134	0.216	< .001
	No Prov.	105	134	0.784	< .001
dominance	Provocation	65	134	0.485	0.796
	No Prov.	69	134	0.515	0.796
masculinity	Provocation	62	134	0.463	0.437
	No Prov.	72	134	0.537	0.437

Note. H_a is proportion ≠ 0.5

Table 6

Manipulation Check Means (SDs)of the CRTT in Study 1.

Condition	Anger towards Opponent	Likability of Opponent	Noise Level Punishment
provocation	2.90 (1.14)	2.38 (0.91)	2.78 (0.99)
no provocation	1.41 (0.77)	3.32 (1.04)	1.97 (1.03)

SAM ratings were analyzed with separate 2 (*Condition*: provocation, non-provocation) × 2 (*Time*: pre, post CRTT) mixed ANOVAs for each affective state measure (pleasantness, arousal, dominance). The ANOVA of pleasantness revealed a significant main effect of *Time*, $F(1,81) = 11.40$, $p = .001$, $\eta_p^2 = .123$, and an interaction between *Condition* and *Time*, $F(1,81) = 7.71$, $p = .007$, $\eta_p^2 = .087$. Participants in both conditions felt worse after the CRTT, and even more so in the provocation condition. The ANOVA of arousal only revealed a main effect of *Time*, $F(1,81) = 4\text{-}37$, p = .040, $\eta_p^2 = .051$. Participants felt more aroused after the CRTT, irrespective of provocation. The ANOVA of dominance ratings yielded a main effect of *Time*, $F(1,81) = 23.32$, $p < .001$, $\eta_p^2 = .224$, with participants feeling less dominant following the CRTT procedure, irrespective of condition. ANOVAs with the between subjects factor *Condition*, did not result in significant differences in arousal, $F(1,81) = 2.66$, $p = .107$, $\eta_p^2 = .032$, or dominance, $F(1,81) = 1.39$, $p = .242$, $\eta_p^2 = .017$.

4.2.3. Interim Discussion

In support of the research hypothesis, the aggressive behavior of another person in the CRTT had a lasting influence on how the other's face was remembered: The face of the aggressively behaving opponent was reconstructed with features that made him look less trustworthy and less likable in comparison to the non-aggressively behaving opponent. Importantly, latter judgements were obtained from a naïve sample of participants who had no knowledge of the aggressive interactions. This finding suggests consistent changes in the visual representation of the interaction partner, which can lead to differences in social judgements based on the impression of the face alone. A plausible candidate for this visual change could be the recollection of a facial displays that are associated with high states of aggressiveness and hostility (e.g., anger displays). Contrary to our hypothesis, however, we did not observe differences in masculinity and dominance, suggesting that facial traits associated with these dimensions (e.g., width to height ratio, brow protrusion) are less susceptible to a memory distortion following aggressive interactions.

4.3 Study 2

Aggressive social interactions, and associated social perceptions, are not only influenced by the social behavior of the interaction partners but also by the reactions of the interaction partner to one's own actions. In line with this interactive perspective, laboratory studies demonstrated that revenge seeking is influenced by opponent displays of pain or anger to (retaliatory) physical punishments. More specifically, it was found that expression of pain after a retaliatory sound punishment is evaluated positively by revenge-seeking people (Eder et al., 2020; Mitschke & Eder, 2021) and reduces revenge seeking in the next trial (Eder et al., 2020). Study 2 therefore investigated whether the emotional reaction of the opponent becomes integrated into the mental representation of the opponent's face. To this aim, participants could additionally observe the emotional reaction of the opponent who reacted to the sound blasting either with a display of pain or with a display of anger (group design). It was hypothesized that expressions of pain after a sound punishment should cause a more favorable representations of the opponent in comparison to the condition in which the same opponent reacted with anger. In short, the naïve sample of participants should judge CIs of opponents who reacted with pain as more trustworthy, less hostile, and less dominant compared to opponents who expressed anger after a sound punishment.

4.3.1. Method

Participants

For the image-creation phase, we recruited 80 participants (56 female, M_{age} = 26.08, SD = 8.28) via the subject management pool of our university. Participants received 8€ for the study participation. The opponent always made aggressive sound blast choices during the CRTT; depending on the experimental condition, however, he reacted differently to participants' sound punishments with facial displays of pain or anger. Assignment to the experimental conditions (pain: N = 40; anger: N = 40) was random. Sample size planning was based on the statistical power considerations described for Study 1.

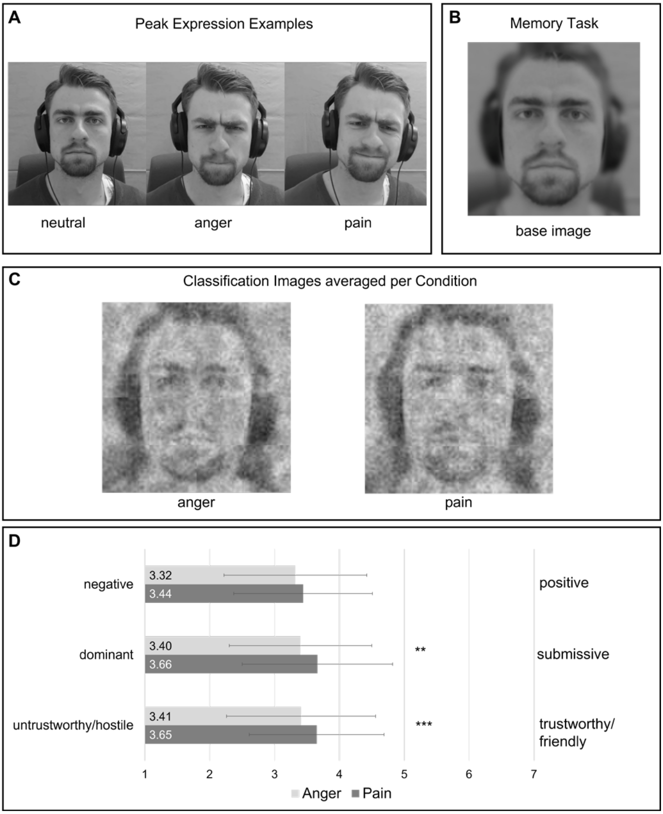
The image-rating sample (Sample 2) was recruited via Prolific and consisted of 295 participants (157 female, 124 male, 2 divers; M_{age} = 35.68, SD_{age} = 14.06; demographic information for 12 participants is missing). Each participant was paid 1£. Sample size estimation for Sample 2 was based on Study 1. Since Study 1 tested differences between extreme groups (highly provoking versus non-provoking) we expected smaller effect sizes for Study 2. A pre-registered a-priori power analysis (performed with G*Power3) for a one-sided dependent sample t-test with power = .80 and an estimated effect size dz = 0.15 yielded a sample of 277. In anticipation of potential data dropouts, we recruited 300 participants. Five participants were excluded from the analysis due to incomplete data or failed attention checks. A post hoc sensitivity analysis with the obtained sample of 295 participants and 0.80 power yielded a detectable effect size minimum of $d_z = 0.16$.

Materials

Videos showing emotional reactions of the opponent to sound blasts were taken from in-house created research materials (Eder et al., 2020). The videos were 3s long and showed a Caucasian white male wearing headphones who reacted with either a neutral, anger or pain expression (please see Figure 10 for exemplary images). In a norming study (Eder et al., 2020), the photograph of the videotaped male model featuring a neutral facial expression yielded ratings of attractiveness (*M* = 2.90, , *SD* = 1.11), masculinity (*M* = 3.53, *SD* = 0.96), and aggressiveness (*M* = 3.83, *SD* = 0.96) in the medium range (1 = not at all, 5 = extremely). The anger emotion expression videos of the opponent yielded a mean rating of 4.36 (*SD* = 1.01) on the dimension anger, mean index of remaining negative categories (sadness, pain, disgust, fear) = 2.08, and the pain emotion expression videos received a mean rating of 3.35 (*SD* = 1.09) on the dimension pain, mean of other negative categories = 2.09 (range: 1 = not at all, 5 = extremely). Noise image creation and CI-construction procedures were identical to those used for Study 1.

Figure 10

Materials, Classification Images, and Ratings for each Condition in Study 2



Note. Panel A shows screenshots taken from the video materials during the peak expression. Panel B depicts the base image. Panel C shows the classification images. Panel D depicts mean ratings of each CI on rating scales, with the score 4 indicating indifference. Error bars show *SD*s. ** $p < .01$.

Procedure

The CRTT was identical to Study 1 except for the inclusion of opponent reactions to (retaliative) noise punishments and fixed win/loss sequences. In addition, all opponents now made aggressive (level 3-5) sound blast choices for provocation of revenge seeking. In win trials, participants watched a short video feed (without audio) of the opponent reacting to the noise punishment either with a facial expression of anger or with expressions of pain (between-design). However, anger and pain reactions were only shown in four out of twelve win trials (trial numbers 9; 14; 18; 24) in order to enhance the plausibility and saliency of the emotional reaction; in the remaining win trials, the opponent showed no distinctive reaction (i.e., a calm face expression) (for a similar procedure see Eder et al, 2020; Mitschke & Eder, 2021). Order and frequencies of anger and pain reactions, as well as levels of sound punishments chosen by the fictitious opponent, were identical in both experimental groups.

After the CRTT task (25 trials), participants were asked to indicate how likable the opponent was and how angry they felt towards the opponent. In the subsequent memorization phase, participants viewed a greyscale image of the opponent (with a neutral face expression) for 2 minutes, during which they should memorize the face. Then, participants completed several questionnaires: four items about justice satisfaction and deservingness (adapted from Funk et al., 2014b), the German version of the Interpersonal Reactivity Index (IRI; Davis, 1983), the so called “Saarbrücker Persönlichkeitsfragebogen SPF (IRI) zur Messung von Empathie” (Paulus, 2009), and six items of the Aggressive Motives Scale (Anderson & Murphy, 2003). Items of justice satisfaction and Aggressive Motives can be found in the Appendix B. The image classification task was identical to that used in Study 1 with the exception that the number of trials was 500.

Image rating phase. The CI images generated by the participants were averaged for each emotion condition and subsequently rated by Sample 2. As in Study 1 the image rating phase started with a display of all CIs in a grid (30 s), followed by a separate display of each individual CI for rating. Rating scales ranged from 1 “very untrustworthy” to 7 “very trustworthy” (trustworthiness), 1 “very hostile” to 7 “very friendly” (hostility), and from 1 “very dominant” to “7 “very submissive” (dominance). For a global evaluation of the facial expression, we added the item “How would you classify the expression of the face. Does it look more negative or positive to you?”, ranging from 1 “very negative” to 7 “very positive”. Images were shown in random order.

4.3.2 Results

An item correlation analysis (as described in Study 1) was conducted. An index was computed for the items assessing untrustworthiness and hostility (item correlation $r(133) = .71$, $p < .001$), we will refer to this dimension as the trust index from now on.

Two dependent t-tests were carried out to compare each judgment dimension (trust index, dominance) between images created based on pain and anger expressing opponents and tested against a Bonferroni-adjusted alpha value of .025 (0.05/2).

Image Ratings (Sample 2).

The dependent sample *t*-tests showed significant differences in the trust index, (comprised out of trustworthiness and friendliness) and dominance (see Figure 10). The CI based on the opponent expressing pain was rated as less untrustworthy and hostile (trust index), $t(294) = 3.42$, $p < .001$, $d_z = .200$, and less dominant, $t(294) = 3.15$, $p = .002$, $d_z = .183$, in comparison to the CI based on opponents expressing anger. The difference in the global rating of the affect direction (positive/negative) of the facial expression was not significant, $t(294) = 1.70$, $p = .091$, $d_z = .099$. See Table 7 for the descriptive values. As in Study 1, we also conducted an exploratory direct comparison rating task. Results pertaining the direct image comparison can be found in the Appendix B.

Table 7

Classification Image Rating Means (*SDs*) in Study 2

CI Type	trust (index)	dominance	expression valence
pain	3.65 (1.04)	3.66 (1.16)	3.44 (1.07)
anger	3.41 (1.15)	3.40 (1.10)	3.32 (1.10)

Manipulation Checks of Provocation (Sample 1). In line with the research of Eder and colleagues (2020), difference scores were calculated via subtracting the chosen sound volume after win trials with anger/pain reactions from the chosen volume in the preceding trial. A negative subtraction score indicates a reduction of the selected punishment intensity, whereas a positive score indicates an increase. Comparison against zero showed that participants significantly reduced the level of desired punishment after the observation of a pain expression, $M = -0.32$, $SD = 1.10$, $t(39) = -3.41$, $p = .002$, $d = 0.54$. Anger expressions on the other hand, produced no significant change, $M = -0.09$, $SD = 0.97$, $t(39) = -1.03$, $p = .308$, $d = 0.16$. A direct comparison of mean difference scores between the anger and pain condition produced no significant difference, $t(78) = -1.83$, $p = .071$, $d = 0.41$.

Conditions did not differ in ratings of anger towards the opponent, $M_{pain} = 3.25$ ($SD = 1.39$), $M_{anger} = 3.15$ ($SD = 1.17$),

$t(78) = 0.35$, $p = .729$, $d = 0.08$, or likability, $M_{pain} = 2.48$ ($SD = 1.01$), $M_{anger} = 2.20$ ($SD = 0.91$), $t(78) = 1.28$, $p = .205$, $d = 0.29$.

Participants in the anger and pain conditions did not differ significantly in their self-reported empathic abilities, $t(78) = 0.88$, $p = .383$, $d = 0.20$, justice satisfaction, $t(78) = 0.17$, $p = .867$, $d = 0.04$, or aggressive motives, all ts ≤ 1.67, all ps $\geq .099$, all ds ≤ 0.37 . For descriptive values of these measures see Appendix B.

We conducted a 2 (*Condition*: pain, anger) × 2 (*Time*: pre, post CRTT) mixed ANOVA of each of the affective state measures (pleasantness, arousal, dominance). After the CRTT, participants indicated more negative affect, main effect *Time*, $F(1,78) = 39.90$, $p < .001$, $\eta_p^2 = .338$; higher arousal, $F(1,78) = 16.24$, $p < .001$, $\eta_p^2 = .172$ and less feelings of dominance, $F(1,78) = 32.42$, $p < .001$, $\eta_p^2 = .294$. The ANOVA did not yield a significant main effect of condition for ratings of pleasantness, $F(1,78) = 0.17$, $p = .682$, $\eta_p^2 = .002$, arousal, $F(1,78) = 1.32$, $p = .254$, $\eta_p^2 = .017$, or dominance, $F(1,78) = 0.09$, $p = .760$, $\eta_p^2 = .001$. The analysis did also not result in any significant interactions of *Condition* and *Time*, for the dimensions pleasantness, $F(1,78) = 0.42$, $p = .521$, $\eta_p^2 = .005$, arousal, $F(1,78) = 0.01$, $p = .903$, $\eta_p^2 < .001$, or dominance, $F(1,78) = 0.01$, $p = .928$, $\eta_p^2 < .001$. For descriptive statistics see Table 8.

Table 8

Self-Reported Affect Means (*SDs*)

Condition		Valence	Arousal	Dominance
pain	pre	3.95 (0.71)	2.38 (0.95)	3.95 (0.93)
	post	3.28 (0.93)	2.78 (0.97)	3.18 (1.11)
anger	pre	3.95 (0.71)	2.15 (0.89)	3.90 (0.84)
	post	3.40 (0.84)	2. 58 (0.96)	3.10 (1.28)

4.3.3 Interim Discussion

As in Study 1, visual representations of aggressive opponents were judged by an independent sample as untrustworthy, hostile, and dominant. Furthermore, the image of the pain-expressing opponent was rated as less untrustworthy, less hostile, and less dominant, compared to the image of the anger-expressing opponent. In the global evaluation measure, both CIs were rated to depict the same degree of negativity within their facial expressions. Overall, these findings demonstrate that perceived emotional reactions to (retaliatory) punishments influence the visual representation of the opponent in addition to the aggressive behavior during the social interaction.

4.4 General Discussion

Social processing of another person's actions in a social interaction influences how this person is represented in memory, including the visual representation of her face. Results from both studies indicate that the experience of aggressive conflict during a competitive interaction resulted in a more negative representation of the opponent's face. Provocation during the interaction led to a pictorial representation of the opponent's face that was judged as less trustworthy and less likable by an independent sample of participants, indicating that perceived traits were inferred from the visual information within the face alone. The second study showed that the representation was also affected by the opponent's emotional reaction to (retaliatory) sound punishments, with suffering opponents being represented more favorable than angry opponents. These findings suggest that social actions and their outcomes affect how a person becomes visually represented in the cognitive system.

Evidence from the perception of faces suggests that visual representations of faces could be biased by the mood of the perceiver (Schmid et al., 2011), social information about the opponent (Mattarozzi et al., 2019), as well as biases in the visual recall, for example due to a higher saliency of negative facial cues (Markovic et al., 2014). Opponents in Study 1 did not show emotional expressions or movements (all images were static), nevertheless opponent representations varied in a systematic way, suggesting that the opponent representation is not solely based on the recall of facial features but may integrate additional social information, such as the opponent state (negative), or affective information, such as the perceivers' state, into memory. This integration could potentially be carried out via adjusting changeable facial features such as emotional expressions. Experiencing a highly aggressive opponent may have resulted in a representation integrating anger expression features (such as brow furrowing). Both expressions portrayed in Study 2 featured similar muscle movements (clenching of the jaw, furrowed brows) and visual features, but resulted in differently judged images, which also favors the assumption that representations were not build on visual recall alone.

If expressive features of opponent representations are potentially altered by social experience, it remains an important question if these alterations are solely valence-based or specific for distinct emotional expressions and the subsequent evaluations thereof. Study 2 did not indicate differences between the overall negativity of the facial expressions portrayed by each represented face, but systematic differences in the judgement of hostility, trustworthiness, and dominance. Specifically, a more favorable representation of opponents suffering from (retaliatory) punishments is in line with previous research that observed pleasant affective reactions in the observer (Mitschke & Eder, 2021) and appeasing effects of pain expressions on retaliatory motivation (Eder et al., 2020). Signs of suffering also decreased the intensity of retaliatory punishments, relative to anger expressions, in the present study, replicating the previous study result of Eder and colleagues (2020).

A biased representation of changeable face features could also account for the observed changes in perceived trustworthiness. Previous research suggested that trustworthiness, inferred from neutral faces, is largely determined by the perceived resemblance to emotional expressions (Oosterhof & Todorov, 2008). For example, Oosterhof & Todorov (2009) reported that persons with an angry face expression were rated as less trustworthy compared to smiling persons.

Keeping invariant facial features mostly constant in memory could aid identity recall and combined with shifts in changeable facial expressions, preserve both, identity and social information about the aggressive nature of the opponent. Models of face perception (e.g. Bruce & Young, 1986; Haxby et al., 2000) typically distinguish between the processing of invariant facial features that aid identification of the person and changeable features of the face (such as expression, eye and lip movement) that facilitate social communication. Although each processing path recruits functionally independent routes, subserved by anatomically distinct brain regions (Duchaine & Yovel, 2015), research suggests a cross talk between these processing routes (Calder & Young, 2005).

Distortions in the perception of invariant facial features were suggested by a study of Balas and colleagues (2015), who reported a change in the remembered fWHR after a competitive interaction. It should be noted, however, that fWHR is also influenced by dynamic facial configurations, such as facial expression of anger, which are correlated with higher perceived fWHR (Merlhiot et al., 2021). While in Study 1, the interaction with a revenge-provoking opponent in the absence of emotional outcomes did not influence ratings of dominance, the anger-expressing opponent in Study 2 was rated as more dominant. This fits with previous research that demonstrated correlations between (male) expressions of anger and judgments of dominance (Hess et al., 2005).

Several questions remain to be addressed in future research. Interacting with the highly aggressive opponent in Study 1 not only affected the social relationship between the interaction partners, but was also associated with more intense sound punishments (i.e., more pain), which by itself could have caused more negative feelings. In fact, the group that interacted with the aggressive opponent reported more unpleasant feelings following the interaction, which could have been caused by the inappropriateness of the opponent's aggressive behavior, the intensity of the aversive sound blasts, or both. Negative mood has been shown to influence face perception, resulting in less favorable impressions of persons (Forgas & Bower, 1987). It remains to be investigated how these different sources of negative affect during the interaction affect person representation. For a differentiation, future studies could manipulate aggressiveness of social interactions without variation of physical pain levels.

Furthermore, only male opponents were presented during the CRTT, which limits the generalizability of the results. Previous studies demonstrated that females are generally perceived as more trustworthy than males (Buchan et al., 2008). In addition, female expressions of anger are typically categorized slower, less accurately, and more unfavorably (Becker, 2017; Dong et al., 2018). The interaction between gender and emotional expressions, and the association between masculine facial features and inferences of aggressiveness, warrant more investigation.

Another interesting question concerns individual differences in the visual representations of opponent. By aggregation of CIs, individual differences in generated images were attenuated to general trends. Since trait aggression has been linked to hostility biases in the perception of faces (Smeijers et al., 2017), future studies could link a perceptual bias of highly aggressive individuals to a more hostile representation of interaction partners.

More recent advances in the methodology of the creation of classification images suggest that image creation based on averaging larger groups instead of subsets of smaller groups of participants, may inflate type 1 error in underestimating the individual variability of each participants mental representation during the creation process (Cone et al., 2020). Hence, the representation data provided in this article should be viewed as central tendencies and future studies should integrate additional image creation methods that take the individual variance stronger into account.

Finally, it would be worthwhile to investigate the role of (more) hostile visual representations for the instigation and maintenance of aggressive behavior. Perceptions of untrustworthiness could foster further aggressiveness, since persons with less trusted faces have been shown to be punished more intensely (Wilson & Rule, 2015). The integration of low trustworthiness into the facial representations of opponents could also cause less personal concern for opponents, since trustworthiness has been shown to modulate empathic responses (Sessa & Meconi, 2015).

Conclusion

In summary, the present research indicates that competitive aggressive interactions modulate the visual representation of opponents, resulting in less favorable visual representations. While representations of opponents were rated generally unfavorably, presumably due to the punitive-competitive interaction context, aggressive behavior of the opponent resulted in visual representations that were even more negative. The mental representation of the opponent was also influenced by dynamic interaction features such as the emotional reaction of the opponent to (retaliatory) sound punishments. In comparison to anger, pain expression after punishments resulted in more favorable representations of the opponent. These findings demonstrate a joint influence of social context, social behaviors, and observed emotional display on the formation of a visual representation of an interaction partner's face.

CHAPTER V

Chapter 5 - General Discussion

5.1 Summary of Findings

All presented research lines support the assumption that the observation of opponent states following punishment (win trials) modulates subsequent punishment. The opponent state is evaluated not only broadly according to valence, meaning punishment outcomes are not solely assessed based on the direction of valence, but distinct displays of negative emotions result in different adjustments in punishment. Observing opponents reacting with anger or sadness led to no adjustments in punishment choices, indexed by non-significant changes from the previous trial to the trial following the opponent reaction. Pain reactions were consistently across all studies followed by significant decreases in punishment, replicating the influence of the observation of pain cues observed by Eder and colleagues (2020).

Opponent displays of joy (or schadenfreude, Chapter 3) following the punishment of the participant (loss trials) were followed by significant increases in punishment. This result highlights the constant monitoring of the opponent during the interaction, even in negative outcome trials (losses), and its potential for the adjustment of punishment behavior in the form of added provocation, potentially increasing the desire for revenge. This increase in revenge desires could potentially lead to an adjustment of desired opponent states, increasing the overall motivation to aggress against the opponent to change their state for the worse.

Results obtained in Chapter 4 indicate that mental visual representations of the opponent were affected by provocation level, resulting in less favorable opponent representations. The changes in visual representation were indexed by social judgements of the opponent faces created from memory by an unrelated sample of participants, who judged the created face representations free from context. Provocation led to representations that were judged as less trustworthy and likable, mirroring the untrustworthy high punishment behavior experienced by those who created the face images.

Additionally to representations being sensitive towards experienced punishment behavior (provocation), an additional experiment showcased that opponent punishment reactions of anger or pain, although not seen often (only in 4 out of 12 trials), significantly influenced the representations of the opponent face as well, further highlighting the importance of opponent monitoring and subsequent memory adjustments according to the experienced opponent state. The reduction of punishment levels following pain displays coincided with more favorable opponent representations, indexed by higher trustworthiness and submissiveness when compared to the images created following state indicators of target anger. The observation of opponent pain hence does not only have consequences for the adjustment of punishment, but also influences the memory of the opponents' face.

In summary, all reported results are in line with the assumption of the *comparable suffering* hypothesis (Frijda, 2007), that the experience of suffering of the target, fulfills revenge desires and hence decreases aggressive motivation. The studies reported can however do not allow direct conclusions about the comparability aspect of suffering. The comparable suffering hypothesis states that the affective suffering of the original offender should be equal or similar to the suffering experienced by the avenger due to the original offense. Concerning the "comparable" aspect of the comparable suffering hypothesis, the reported studies cannot disentangle the relationship between the amount of self-experienced suffering and opponent suffering. Firstly, none of the reported studies measured the experience of initial negative affect due to the provoking behavior in itself. Hence, we cannot compare the experienced suffering with the amount of suffering the opponent displays. Second, when comparing the choices of noise levels between the participant and provoking opponent, participants chose on average lower punishments than the provoking opponent. It remains to be investigated how the directly experienced suffering by the individual modulates their expectancy for opponent suffering in additional studies.

All studies were designed with the assumption that harm serves as a proximate goal of revenge, in line with the definition of aggressive behavior (DeWall et al., 2013). As for more ultimate goals of revenge, the following framework proposes the integration of proximate outcome appraisal as a first step to allow for more precise assumptions about the mechanisms that guide action choice during vengeful interactions.

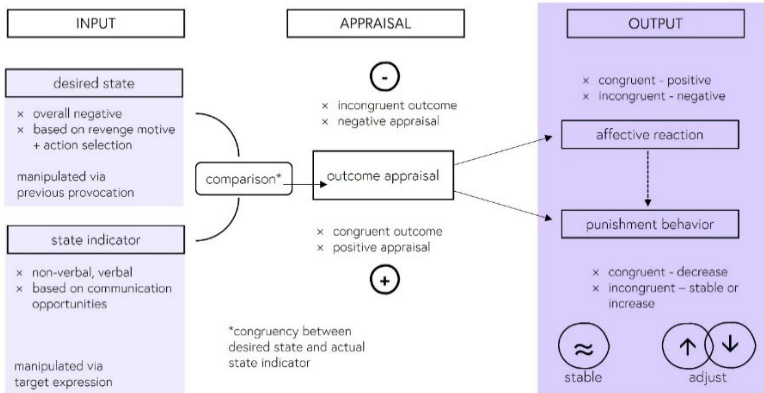
5.2 Integration into Cybernetic Framework

To formalize the proximate processes involved in the appraisal of punishment outcomes and to allow for a discussion of the potential modulating properties of more distal processes, such as empathy or mood regulation efforts, the following paragraphs will provide a framework outlining the relationship between desired outcomes (revenge motivation), opponent states (observed environmental changes) and the subsequent affective and behavioral responses within a revenge episode (for a graphical overview please see Figure 11).

The proposed processes and structure are based on a cybernetic approach to action control (Carver, 2006), building on the assumption that opponent monitoring serves as an information input that is taken into account during the next action selection phase, resulting in a potential adjustment of the subsequent punishment level. This approach adds the assumption that a constant monitoring process of the target during aggressive interactions is necessary to inform the individual of the current state of the interaction and whether to adjust subsequent behavior.

Figure 11

Cybernetic Framework: Proximate Outcome Appraisal



During all sets of studies, we manipulated target state indicators and desired states (both highlighted in lavender), while measuring the subsequent outputs in the form of action choices (punishment choice) and affective reactions during the observation of opponent state indicators (highlighted in purple). This approach allows for more precise predictions of different punishment outcomes based on the interaction between revenge motivation (desired target states) and observed target state (target state indicator). Appraisals of the punishment outcome are hypothesized to be based on the comparison between the desired opponent state, which was manipulated via prior provocation and the actually perceived opponent state indicators, manipulated via emotional outcome reactions.

In the following paragraphs, each framework part will be discussed based on the empirical results obtained in Chapters 2 to 4 and extended to demonstrate how the framework may be adjusted to investigate more distal processes that influence each component.

5.2.1 The Desired Opponent State

The desired state of the opponent, the target goal of the interaction, is theorized to be formed by the quality and intensity of the revenge desire. In the reported sets of studies, provocation was utilized to increase anger (as indicated in all studies) towards the opponent, and due to higher noise punishment levels in the provocation condition, create a more intensive aversive experience (indicated via higher ratings of arousal and negative valence in all studies). All of the aforementioned factors should lead to an increase in revenge desire. Changes in the desired outcome state should in turn lead to different appraisals of the outcome. These differences can be assessed via comparing the high to no provocation condition. Observations of opponent suffering resulted in different affective perceiver reactions based on prior provocation, with positive reactions towards highly provoking suffering opponents and negative reactions towards the suffering of non-provoking opponents.

5.2.2 Appraisal of Punishment Outcomes

Before a second punishment decision is made, the individual has to appraise the outcome of the current punishment decision. During this stage, the avenger monitors the opponent to detect signs of the current opponent state. This evaluation is carried out under an aversive state, with high levels of anger, due the prior provocation and should favor the detection of signs of potential threat (Baumann & DeSteno, 2010) to inform the assessment of potential harmful consequences of punishment decisions. Neuroscientific studies correlate the punishment decision phase with activations of mentalizing networks (Chester & DeWall, 2016; Lotze et al., 2007), theorizing that the observed activations are due to comparing personal goals and states with the state of the opponent.

Target state indicators may consist of any observable input that informs the individual about the current state of the target. As reviewed in Chapter 1, target states may be inferred from verbal or written statements, as in Denzler et al. (2009), emotion indications via drawings, as in Arriaga et al. (2019), scale-based reporting of shock or noise intensity towards the target, as in Chester et al. (2016) or emotional facial expressions, as in Eder et al. (2020) and studies reported in Chapter 2 to 4. Different modalities of state indicators may require varying amounts of effort to decode and appraise during the interaction.

The opponent state in our case indexed via facial expressions, is compared to the desired state, to indicate whether the desired goal of the interaction is reached. The closer the target state to the proximate goal of harm, the lesser the subsequent punishment. Observing target states that are negative and therefore congruent but not identical with the goal (such as anger or sadness) stabilizes subsequent punishment behavior, as also observed by Eder et al. (2020). Only cues directly signaling pain and therefore the successful outcome signaling harm, resulted in a downregulation of punishment.

Sadness facial expressions were chosen as a visual display of submissiveness, which was interestingly not followed by significant decreases in punishment nor distinct facial muscle reactions, replicating Eder et al. (2020). This does not indicate that sadness outcomes never modulate revenge behavior but could be an interesting addition to the investigation of different punishment motives, such as deterrence, since submission should signal willingness to withdrawal from the interaction. In the case of the used CRTT paradigm, withdrawal might not have been a likely interpretation of opponent behavior since all blocks were designed to have multiple rounds and each opponent was encountered twice, not allowing for a withdrawal from the competition. Interestingly, sadness expressions did not lead to a positive affective reaction (as indexed by facial muscle responses), which further fosters the assumption that, at least in the chosen design, submission was not a proximate goal of the punishment behavior.

Although sadness could be perceived as an indication emotional of harm (Horstmann, 2003), the perception of sadness in the opponent does not seem to fulfill the goal of harming the opponent, indicated by the absence of a positive response and the absence of a downregulation of punishment. Going back to the comparability aspect of the comparable suffering hypothesis, this could be interpreted as a lack of congruency between the perceivers' punishment experience (the aversive noise) and the opponent state, potentially indicating that sadness reactions are not perceived as comparable states to physical punishment.

Opponent monitoring is presumed to take place over the whole interaction, not only during the assessment of punishment outcomes. The observation of cues signaling positive target states, which are largely incongruent to the desired opponent state, are shown to increase provocation levels, resulting in an up-regulation of punishment, as observed in Chapter 3. The observation of positive target states was only interpreted as schadenfreude and attributed to a highly negative attitude when paired with previous provocation. Additionally, the observation of neutral target states during the provocation phase (following losses), lead to negative affective reactions, which is in line with the assumption that the image of the highly provoking opponent acted as a negative stimulus in itself. These results again highlight the importance of the social context in which the target state indicators are observed.

5.2.3 Subsequent Punishment Selection

Goal progress (indexed via negative opponent state indicators) motivates further goal directed actions (Koo & Fishbach, 2014), which should lead to stable punishment choices, indexing that the individuals desire to take revenge has not yet been satisfied. A decrease in punishment following pain displays could be interpreted as behavioral consequence of goal attainment, which should in turn downregulate further aggressive actions.

Results from all study lines varied punishment outcomes experimentally which was followed by behavioral change in the case of pain cues being present. To investigate the causal relationship between positive affect felt during the pain observation and the regulation of the subsequent punishment decision, additional studies and analyses will be needed. It remains to be empirically investigated how the affect experienced during the opponent observation informs the subsequent decision phase resulting in adjustments of punishment. The presented framework, modeled after cybernetic control processes, merely presents a theoretical integration that seems plausible, but remains to be empirically tested.

Additionally to monitoring goal progress in comparing desired and observed opponent states, the subsequent action selection is potentially influenced by reward expectancies and potential conflicts between the desire to harm and learned norms and prosocial tendencies. Action selection during high provocation involves the activation of reward processing-based structures (Bertsch et al., 2020). During the decision phase, social standards and norms should also be activated, resulting in increases in conflict based processing, due to the incongruence between reward expectancy due to goal achievement and social norm appropriate behavior options (Lotze et al., 2007; Repple et al., 2017). Greater connectivity between the ventral lateral prefrontal cortex (VLPFC) a region implicated in the inhibition of reward-based processing, as well as aggressive impulses (Mehta & Beer, 2010) and the NAcc, has been shown to correlate with lower intensity punishments, indicating a potential inhibitory process that down-regulates the reward estimation of potential aggressive actions during the decision phase (Chester, 2017).

The appraisal of opponent outcomes not only regulates subsequent punishment behavior, but can also result in distortions in the representation of opponents, as reported in Chapter 4. Results presented in Chapter 4 indicate that punishment outcomes not only modulate the course of future punishment, but also have more long-term consequences for the representation of the previously encountered opponent. Representations of entities (such as people or objects) are influenced by multiple information sources, such as the situations in which these were encountered, bottom-up visual features or stereotypic expectancies (Freeman & Ambady, 2011). In the case of aggressive competition, the increased negative affect, as well as observed negative behavior, led to less favorable mental representations of the opponents' face, fusing the experienced affect and behavior into a coherent negative representation of the person. The more negative representation of the opponent may affect subsequent interactions since expectations of untrustworthiness could favor future aggressive actions towards the mistrusted person. Persons with less trusted faces have been shown to be punished more intensely (Wilson & Rule, 2015) and empathic reactions towards untrustworthy faces in pain differ on a neuronal level (Sessa & Meconi, 2015), with untrustworthy faces evoking lesser empathic processing, even in the absence of explicit information about the other person. The representation of opponent faces in memory as untrustworthy and unlikeable may correspond with a subsequent avoidance of future affiliative actions (Krumhuber et al., 2007), fostering the avoidance of empathy and compassion towards the opponent further.

Additionally, the changes in opponent representation were not only affected by the experience of provoking behavior but also affected by the outcomes of punishment, as indicated by emotional target reactions. The more favorable representation of opponents expressing pain strengthens the importance of outcomes indicators for subsequent interactions and showcase that the observation of opponent suffering is not only positively appraised but also integrated into the face memory of the opponent.

5.3 Revenge and Emotion Regulation

The positive affective reaction present during the observation of opponent pain is not only a potential consequence of goal approach but can also be discussed as a function of aggressive behavior in itself. As discussed in Chapter 1, revenge interactions are often preceded by an aversive and negative state within the provoked individual, accompanied by the experience of anger (Baron, 1974; Ramírez & Andreu, 2006). The presence of negative emotions, such as anger, should foster regulation processes to relief the experienced negative state, resulting in approach behavior towards rewarding experiences, relieving negative affect (R. J. Larsen, 2000; Tice & Bratslavsky, 2000). The notion that individuals strive for homeostatic balance via self-regulation has also been proposed in recent models of emotion regulation (Tamir, 2021). Vengeful behavior has been connected to the regulation of mood following provocation (Gollwitzer & Bushman, 2012). Given that the state of the individual is already negative (due to provocation and arising feelings of anger), the individual may prioritize behavioral choices that improve their mood, or that they believe will improve their mood, especially since focusing on negative states and their emotional regulation has been shown to lower self-regulatory resources in other domains (Tice & Bratslavsky, 2000).

Although mood repair has been shown to not be an ultimate goal (Gollwitzer & Bushman, 2012) and a net positive mood is not observed as a consequence of revenge actions (Chester et al., 2021; Eder et al., 2020, 2021; Mitschke & Eder, 2021), research has indicated modulations in aggressive tendencies based on anticipated mood repair (Bushman et al., 2001) and its modulating properties during aggressive responding following rejection (Chester & DeWall, 2017). The idea of mood repair as a consequence of successful revenge seems to be deeply rooted in human beliefs (Bushman, 2002), but could be hindered by rumination (Eadeh et al., 2017) and the impact of the previous negative affect due to provocation.

Even when recruited under the mechanism of mood repair, anger about the original offense may still be a salient reference point for the self-report of affect. The anger experienced during the overall procedure may overshadow a measurement of mood repair in the form of short relieve. All participants reported negative feelings and anger towards the opponent irrespective of the punishment outcome. These self-reports were obtained after multiple interactions, between game blocks and the absence of systematic variance in these results further indicates that it remains difficult to measure affect following punishment outcomes on a self-report level.

Even if punishment is sought after as a means of anticipated mood repair, the expected mood change may not occur since the magnitude of positive affect may not be large enough to shift the overall affective experience. As also seen in Chapter 3, the overall experience and the attitude towards the opponent stays negative, with previous provocation being integrated into the representation of the opponent as potentially untrustworthy and negative overall.

The presented research lines add to the importance of the investigation of positive affect as a regulator of aggressive responding, but also stresses that it may not be the enactment of revenge in itself that is rewarding, but the effect it creates on our environment. The monitoring of the environment as an indicator of successful action should be integrated into ideas about aggression catharsis, which are often linked to the venting of anger, as a form of arousal release, irrespective of the outcome of the aggressive action (Bushman, 2002; Bushman et al., 2001). The action of choosing the punishment and it being delivered in itself was only followed by positive affect if the punishment outcome signified a salient cue of suffering, aka pain. Empirical studies on the affective nature of aggressive episodes should therefore keep in mind that the affective experience during aggressive actions is constituted not only of emotional consequences from acting out vengeful behavior but are modulated by the immediate effects these behaviors have on its targets, as proposed by the framework of proximate processes during the evaluation of the revenge outcomes discussed earlier.

The potentially rewarding nature of revenge has been implicated to play a pivotal role in the chronification of aggressive behavior (Chester, 2017). Individuals high in aggressive traits have been shown to not only experience higher levels of anger

(Wilkowski & Robinson, 2010), but also more stable levels of positive affect during the punishment of others (Chester et al., 2021). The relief (short-lived positive affect) experienced during the observation of goal attainment could foster positive associations between acting aggressively and favorable outcomes, guiding the behavior selection for the next revenge episode. This process is in line with the idea that chronic forms of aggressive responding might be fostered via reinforcement learning-based processes.

5.4 Aggressive Competition hindering Compassion

As discussed in Chapter 1, the reported studies not only allow for an investigation into revenge motivation mechanisms, but were also designed to shed light on the potential empathy modulating qualities of aggressive interactions. The following paragraphs will integrate the obtained results into recent theories of empathy and discuss potential implications for the link between aggression and empathy.

The knowledge of the suffering of another person and the subsequent experience of compassion has been indicated as a prerequisite to helping behavior and a potential deterrent from harming another person (Eisenberg & Miller, 1987). It is often suggested that emotions evoked by empathy (such as guilt, shame, compassion) inhibit aggressive action tendencies (Eisenberg, 2000; Singer & Klimecki, 2014). Nevertheless, a meta-analysis on the relationship between trait empathy and aggression only reported very small effect sizes, with high levels of trait empathy being mostly correlated to decreases in verbal aggression (Vachon et al., 2014). The authors point out that many studies measure empathy as a trait-based and fairly stable construct, which may fail to address the changes in empathic processing towards disliked targets.

Empathy trait questionnaires are designed to measure a general tendency to engage in empathic thoughts and feelings, but often do not take into account the specific person or situation these thoughts and feelings relate to. The potentially empathy-eliciting target does play a pivotal role in the magnitude of empathic processing, with empathy-related processes being lowered in response to outgroups (Sachisthal et al., 2016), and disliked or unfair others (Likowski et al., 2008; Singer et al., 2006). This so called *empathy gap* is discussed to enable the enjoyment of harm in others (schadenfreude) and may result in counter-empathic reactions, deteriorating interpersonal relations (Cikara et al., 2011; Vanman, 2016). The ease of social modulation of empathic processes, favors the assumption that the concept of empathy can be more fully understood when adding state-based components that are flexible depending on context and motivation.

The *motivational account of empathy* (Zaki, 2014) postulates that empathy can be viewed as a state-based construct modulated by motivation, adding dynamic empathic processes to the trait construct of empathy. Facial mimicry is often treated as an implicit measure for dynamic state based empathy, with a recent meta-analysis indicating that the correlation between trait empathy and mimicry is strongest for empathy sub-facets that evolve emotional empathy (Holland et al., 2021). An absence or attenuation of facial mimicry responses has been shown to be related to negative attitudes towards the interaction partner (Likowski et al., 2008).

Punishment outcome reactions, be it either deserved or undeserved (provoked or unprovoked), never resulted in participant mimicking behavior, indicating a general attenuation of emotional empathy-related processing towards opponents during competition. Only in the case of undeserved pain did we observe a negative affective response pattern, potentially indicative of personal distress or concern towards the opponent. This adds to the assumption that empathy-related processes have to be motivated in the current context, which is in line with evidence indicating lower empathy effort towards competitors. The involvement of corrugator attenuation of mimicry responses during high provocation, as presented in Chapter 3, supports the notion that even mimicry towards highly affiliative gestures such as smiling can be modulated by provocation, even though smiling mimicry responses are highly habitual and difficult to suppress (Korb et al., 2010).

In the realm of the proposed framework, changes in empathy may affect a multitude of processing stages. To allow for more concise predictions on the influence of empathy on aggression, the different sub-facets included in the empathy construct should be reviewed separately (Batanova & Loukas, 2011). Perspective taking and mentalizing, as indicated by previous research (Chester & DeWall, 2016), could be necessary processes to allow for a decision-making process that factors the desired target state into the selection of subsequent revenge actions. Within the proposed framework, a lack of affective perspective-taking abilities, the ability to infer affective states from observing others (Healey & Grossman, 2018), could result in an underestimation of the target state, potentially resulting in higher thresholds to detect suffering to achieve a congruent punishment outcome.

Empathic emotional reactions such as guilt or shame, would potentially hinder goal pursuit in an aggressive interaction and may therefore be weakened during the pursuit of revenge. The regulation of emotions has been theorized to not only occur following the experience of an emotion, but also to alter the process of the occurrence emotions before their unfolding (Gross & Feldman Barrett, 2011). Future studies investigating empathic processing during competition and aggressive interactions should therefore differentiate between the downregulation of aggressive tendencies due to emotions being evoked by empathy and changes in target perception due to a potential lack of affective perspective taking.

5.6 Limitations

All studies presented investigated revenge as reactive aggression, employing a highly controlled experimental environment and fixed time frames. This set up allowed for a controlled manipulation of outcomes, while keeping the punishment options stable and scalable. While the choice of experimental approach comes with the aforementioned benefits, it also has some considerable limitations.

First, there is the absence of a disengagement or no punishment option, limiting the options of action choices and potential opponent state indicators. Many scholars have criticized the use of the CRTT due to its flexible scoring options and lack of non-aggressive response options (Elson et al., 2014; Ferguson et al., 2008; Ferguson & Rueda, 2009). All studies presented followed a new scoring approach (mean change between trials), as introduced by Eder and colleagues (2020), as well as a preregistered analysis plan, which does not allow for scoring flexibility, eliminating one major point of criticism concerning the CRTT paradigm. Nevertheless, they did not employ a completely non-aggressive response option. Studies investigating social behavior following hurtful exclusion suggest that a proportion of participants choses withdrawal from the interaction as a coping mechanism (Peterson et al., 2011; Ren et al., 2018) and studies on revenge suggest that disengagement from the interaction is seen as a valid and often-used form of revenge, especially within personal relationships

(Yoshimura, 2007). None of the presented studies allowed for withdrawal-related actions (except for a complete abortion of the study), or a fully non-aggressive response option. Future approaches should offer a broader scope of actions to be selected to investigate the impact of punishment outcomes on the trajectory of revenge interactions with respect to action selection.

Secondly, the samples utilized in the presented studies were mainly student samples low in trait aggressiveness. Comparisons of populations that are more prone to aggression resulted in significant differences in the upkeep of positive affect during aggressive actions (Chester et al., 2021) and involvement of higher order inhibitory processes during the selection of punishment (Krämer et al., 2008). The enjoyment of pain in others is often associated with abnormal personality traits such as sadism (Chester et al., 2019) and psychopathy (Decety et al., 2013). Study line 1 provides a unique new perspective on the enjoyment of pain, mainly that it is not only reserved for sadists and highly aggressive individuals, but can be observed in normal populations as well. The enjoyment of deserved suffering in opponents can therefore not be considered a uniquely maladaptive reaction, reserved for atypical populations. Nevertheless, the obtained results cannot easily be scaled onto highly aggressive populations. Especially the effects of punishment outcomes on face representation discussed in Chapter 4 could potentially be highly susceptible to hostile attribution biases, which are often correlated with a higher propensity for violence (Anderson et al., 2008). A high level of hostile attribution tendencies may affect the overall intensity of the change in representation following provocation, leading to more pronounced negative representations on the individual perceiver level, allowing for a comparison between individuals based on aggressive motivation and hostility bias and its effects on the representation of interaction partners.

Thirdly, there is the overall gender imbalance in the video materials and participant pool. All results were obtained using male opponents and subsequently male facial cues. Although research on provoked aggression suggests no fundamental gender differences in punishment choices following provocation (Bettencourt & Miller, 1996), the observation and interpretation of state indicators via emotional facial expressions could be influenced by perceiver gender (Arriaga & Aguiar, 2019). Arriaga and colleagues utilized drawings of emotional faces as indicators of the opponent state, providing happiness, sadness, anger, and neutral feedback during a competitive reaction time task and observed gender differences in punishment following sadness and anger. While females reduced punishment following anger or sadness, males did not. Interestingly, empathic responses towards others in pain have been shown to be more pronounced in females (Han et al., 2008), although recent studies indicate that this difference is not due to females being more sensitive at detecting emotional cues (Fischer et al., 2018). Concerning the largely female samples in Chapters 2 and 3, we did not observe pronounced empathetic responses towards provoking opponents, except for the simulation of smiles, and did not replicate the findings of Arriaga et al. (2019). If anything, the largely female samples should theoretically favor negative affective reactions towards suffering opponents (due to potentially increased empathetic processing in females), which we did not observe following provocation. Nevertheless, the interpretation of male target expressions may differ from female targets, as anger is more easily recognized in men (Becker et al., 2007), whereas the expression of sadness in men is often deemed as non-stereotypical (Plant et al., 2000). Future studies should therefore include female targets to further generalize the effects of specific emotion reactions on subsequent aggressive responding with regard to target and observer gender.

5.7 Open Questions and Future Research

5.7.1 Agency and the Hedonic Value of Revenge

The experience of willfully manipulating our environment based on our own actions (self-agency) is a fundamental part of goal pursuit. Feelings of control of the environment can additionally enhance the motivation to further pursue the desired goal (Eitam et al., 2013). Perceptions of agency can vary based on the success of goal attainment (Kumar & Srinivasan, 2014): the more successful the action, the more it is perceived to be high in self-agency (Kip et al., 2021). However, it remains to be tested whether the positive affective consequences of attaining the revenge goal, as observed in the presented studies, require self-agency. As already pointed out by Chester and colleagues (2021), it remains unclear how self-agency modulates the evaluation of punishment outcomes, as most paradigms do not explicitly allow for a comparison of self- and other-inflicted suffering. Concerning the link between action and satisfaction with the outcome, it remains to be investigated if positive reactions towards opponent pain are dependent on self-agency or merely driven by the outcome. The question about the impact of self-agency on the evaluation of punishment outcomes is a uniquely important aspect of third-party punishment and victim satisfaction. Third-party punishments are a common structure in most judicial systems, where justice is carried out via a detached third party. In the case of revenge punishments, a sense of agency concerning the action outcome may not only evoke positive feelings due to goal attainment, but could also give rise to negative affect due to guilt, evoked by the direct responsibility for the observed suffering, as observed in forced-punishment paradigms investigating empathy for pain (Lepron et al., 2014).

5.7.2 Punishment Congruency

All studies presented employed a congruent punishment choice (aversive noise) and outcome state (pain). Physical noise blasts and pain expressions may benefit from a highly congruent fit, whereas sadness expressions may have suffered from less congruency between action and outcome. An incongruency between action and predicted action effect reduces the subjective feeling of agency (Sidarus et al., 2017), potentially resulting in an external attribution of the effect (Blakemore et al., 2002). In the case of opponent reactions, an incongruency between the predicted response towards the chosen punishment, as in the case of the sadness reaction, may lead the perceiver to interpret the opponent's sadness as less related to the punishment and potentially to consider a wider range of external factors, such as the opponent's guilt about their own actions or sadness about losing the encounter.

Additionally to the potential misfit between punishment modality and outcome, the CRTT paradigm in its current form does not allow participants to choose from different forms of punishment modalities. In everyday life, revenge behaviors are selected from a broad range of potential actions (Yoshimura, 2007). Take the example of someone steeling your parking spot on a busy Monday morning. You may initiate revenge desires by feeling unfairly treated and choose to physically aggress via kicking the other car or driver. But you may also choose to roll the car window down and insult the other driver, as a form of verbal aggression. Each of these behavioral options may result in different predicted and desired outcomes, and changes in opponent states and therefore consequently different comparison standards for the appraisal of punishment outcomes. As most aggressive behavioral options include a desire to proximately harm or negatively influence the target person, the investigation of harm cues does not always allow for a differentiation between potential ultimate revenge goals. Future studies should combine a broader range of behavioral choices with additional distinct indications of target states, each matching ultimate revenge goals, such as indications of withdrawal to detect deterrence motives, or as done by Funk et al. (2014), verbal acknowledgements of understanding the reasons for punishment. In combing suffering cues with more complex behavioral choices and outcomes, future studies could shed light not only on ultimate motives of revenge, but also on the importance of congruency between punishment action and outcomes and the interplay between proximate harm and higher order goals.

Concluding Remarks

In treating the provoked individual as seeking a desired state within the offender (such as defeat, pain, compliance) and comparing this state to their observations of the opponent, differences in affect and visual representations of opponents were integrated into a framework of the proximate appraisal of revenge outcomes.

Indications of suffering in the form of pain were shown to reliably decrease subsequent aggressive responding, coinciding with positive affective reactions (Chapter 2), demonstrating the modulation of aggressive responding via opponent state indicators and the importance of positive affect in the progression of vengeful interactions. Affective empathic processes, indexed via negative distress reactions, were attenuated towards opponents (Chapters 2 & 3), potentially easing the approach of the harm goal during aggressive interactions. The visual mental representation of opponent faces was shown to be shifted according to levels of previous provocation and distinct visual representations due to outcome indicators of pain and anger are formed (Chapter 4), providing evidence of the impact of outcome appraisals in the formation of opponent representations, which may aid the avoidance of empathy in future interactions.

Future studies may extend the presented framework to investigate the appraisal of punishment outcomes in highly aggressive individuals, as well as the conditions under which suffering cues different to pain may evoke goal satisfaction, presumably when ultimate revenge goals of deterrence, or understanding are present. The emergence of positive affect during the observation of target pain, which coincides with decreases in punishment, strengthens the importance of positive affect in the regulation of aggressive actions, emphasizing dynamic changes in affect during the pursuit of revenge.

References

Anderson, C. A., & Bushman, B. J. (2002). Human Aggression. *Annual Review of Psychology*, *53*(1), 27–51. https://doi.org/10.1146/annurev.psych.53.100901.135231

Anderson, C. A., Buckley, K. E., & Carnagey, N. L. (2008). Creating Your Own Hostile Environment: A Laboratory Examination of Trait Aggressiveness and the Violence Escalation Cycle. *Personality and Social Psychology Bulletin*, *34*(4), 462–473. https://doi.org/10.1177/0146167207311282

Anderson, C. A., & Murphy, C. R. (2003). Violent video games and aggressive behavior in young women. *Aggressive Behavior*, *29*(5), 423–429. https://doi.org/10.1002/ab.10042

Ansfield, M. E. (2007). Smiling when distressed: When a smile is a frown turned upside down. *Personality and Social Psychology Bulletin*, *33*(6), 763–775. https://doi.org/10.1177/0146167206297398

Arriaga, P., & Aguiar, C. (2019). Gender differences in aggression: The role of displaying facial emotional cues in a competitive situation. *Scandinavian Journal of Psychology*, *60*(5), 421–429. https://doi.org/10.1111/sjop.12568

Ayton, P., Pott, A., & Elwakili, N. (2007). Affective forecasting: Why can't people predict their emotions? *Thinking & Reasoning*, *13*(1), 62–80. https://doi.org/10.1080/13546780600872726

Balas, B., & Thomas, L. E. (2015). Competition makes observers remember faces as more aggressive. *Journal of Experimental Psychology: General*, *144*(4), 711–716. https://doi.org/10.1037/xge0000078

Baron, R. A. (1974). Aggression as a function of victim's pain cues, level of prior anger arousal, and exposure to an aggressive model. *Journal of Personality and Social Psychology*, *29*(1), 117–124. https://doi.org/10.1037/h0035707

Batanova, M. D., & Loukas, A. (2011). Social Anxiety and Aggression in Early Adolescents: Examining the Moderating Roles of Empathic Concern and Perspective Taking. *Journal of Youth and Adolescence*, *40*(11), 1534–1543. https://doi.org/10.1007/s10964-011-9634-x

Baumann, J., & DeSteno, D. (2010). Emotion guided threat detection: Expecting guns where there are none. *Journal of Personality and Social Psychology*, *99*(4), 595–610. https://doi.org/10.1037/a0020665

Baumeister, R. F., Stillwell, A. M., & Heatherton, T. F. (1994). Guilt: An interpersonal approach. *Psychological Bulletin*, *115*(2), 243–267. https://doi.org/10.1037/0033-2909.115.2.243

Baumeister, R. F., Stillwell, A. M., & Wotman, S. R. (1990). Victim and Perpetrator Accounts of Interpersonal Conflict: Autobiographical Narratives About Anger. *Journal of Personality and Social Psychology*, *59*(5), 994–1005.

Becker, D. V. (2017). Facial gender interferes with decisions about facial expressions of anger and happiness. *Journal of Experimental Psychology: General*, *146*(4), 457–463. https://doi.org/10.1037/xge0000279

Becker, D. V., Kenrick, D. T., Neuberg, S. L., Blackwell, K. C., & Smith, D. M. (2007). The confounded nature of angry men and happy women. *Journal of Personality and Social Psychology*, *92*, 179–190. https://doi.org/10.1037/0022-3514.92.2.179

Berkowitz, L. (1974). Some determinants of impulsive aggression: Role of mediated associations with reinforcements for aggression. *Psychological review*, *81*(2), 165–176. https://doi.org/10.1037/h0036094

Berkowitz, L. (1989). Frustration-Aggression Hypothesis: Examination and Reformulation. *Psychological Bulletin*, *106*(1), 59–73. https://doi.org/10.1037/0033-2909.106.1.59

Berkowitz, L., & Harmon-Jones, E. (2004). Toward an understanding of the determinants of anger. *Emotion*, *4*(2), 107–130. https://doi.org/10.1037/1528-3542.4.2.107

Bertsch, K., Florange, J., & Herpertz, S. C. (2020). Understanding Brain Mechanisms of Reactive Aggression. *Current Psychiatry Reports*, *22*(12), 81. https://doi.org/10.1007/s11920-020-01208-6

Bettencourt, B. A., & Miller, N. (1996). Gender differences in aggression as a function of provocation: A meta-analysis. *Psychological bulletin*, *119*(3), 422–447. https://doi.org/10.1037/0033-2909.119.3.422

Blair, R. J. R. (2016). The Neurobiology of Impulsive Aggression. *Journal of Child and Adolescent Psychopharmacology*, *26*(1), 4–9. https://doi.org/10.1089/cap.2015.0088

Blakemore, S.-J., Wolpert, D. M., & Frith, C. D. (2002). Abnormalities in the awareness of action. *Trends in Cognitive Sciences*, *6*(6), 237–242. https://doi.org/10.1016/S1364-6613(02)01907-1

Blocker, H. S., & McIntosh, D. N. (2016). Automaticity of the interpersonal attitude effect on facial mimicry: It takes effort to smile at neutral others but not those we like. *Motivation and Emotion*, *40*(6), 914–922. https://doi.org/10.1007/s11031-016-9581-7

Boecker, L., Likowski, K. U., Pauli, P., & Weyers, P. (2015). The face of schadenfreude: Differentiation of joy and schadenfreude by electromyography. *Cognition and Emotion*, *29*(6), 1117–1125. https://doi.org/10.1080/02699931.2014.966063

Bondü, R. (2018). Is bad intent negligible? Linking victim justice sensitivity, hostile attribution bias, and aggression. *Aggressive Behavior*, *44*(5), 442–450. https://doi.org/10.1002/ab.21764

Boulton, M. J., & Hawker, D. S. (1997). Non-physical forms of bullying among school pupils: A cause for concern. *Health Education*, *97*(2), 61–64. https://doi.org/10.1108/09654289710158393

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

Brinkman, L., Todorov, A., & Dotsch, R. (2017). Visualising mental representations: A primer on noise-based reverse correlation in social psychology. *European Review of Social Psychology*, *28*(1), 333–361. https://doi.org/10.1080/10463283.2017.1381469

Brown, S. L., & Schwartz, G. E. (1980). Relationships between facial electromyography and subjective experience during affective imagery. *Biological Psychology*, *11*(1), 49–62. https://doi.org/10.1016/0301-0511(80)90026-5

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*(3), 305–327. https://doi.org/10.1111/j.2044-8295.1986.tb02199.x

Buchan, N. R., Croson, R. T. A., & Solnick, S. (2008). Trust and gender: An examination of behavior and beliefs in the Investment Game. *Journal of Economic Behavior & Organization*, *68*(3–4), 466–476. https://doi.org/10.1016/J.JEBO.2007.10.006

Bushman, B. J. (2002). Does Venting Anger Feed or Extinguish the Flame? Catharsis, Rumination, Distraction, Anger, and Aggressive Responding. *Personality and Social Psychology Bulletin*, *28*(6), 724–731. https://doi.org/10.1177/0146167202289002

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.

Bushman, B. J., Baumeister, R. F., & Phillips, C. M. (2001). Do people aggress to improve their mood? Catharsis beliefs, affect regulation opportunity, and aggressive responding. *Journal of Personality and Social Psychology*, *81*(1), 17–32. https://doi.org/10.1037/0022-3514.81.1.17

Cacioppo, J. T., Petty, R. E., Losch, M. E., & Kim, H. S. (1986). Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions. *Journal of Personality and Social Psychology*, *50*(2), 260–268. https://doi.org/10.1037/0022-3514.50.2.260

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience*, 6(8), 641-651. https://doi.org/10.1038/nrn1724

Carlsmith, K. M., Wilson, T. D., & Gilbert, D. T. (2008). The paradoxical consequences of revenge. *Journal of Personality and Social Psychology*, *95*(6), 1316–1324. https://doi.org/10.1037/a0012165

Carré, J. M., McCormick, C. M., & Mondloch, C. J. (2009). Facial structure is a reliable cue of aggressive behavior: Research report. *Psychological Science*, *20*(10), 1194–1198. https://doi.org/10.1111/j.1467-9280.2009.02423.x

Carver, C. S. (2006). Approach, avoidance, and the self-regulation of affect and action. *Motivation and Emotion*, *30*(2), 105–110. https://doi.org/10.1007/s11031-006-9044-7

Chang, L. J., Doll, B. B., van 't Wout, M., Frank, M. J., & Sanfey, A. G. (2010). Seeing is believing: Trustworthiness as a dynamic belief. *Cognitive Psychology*, *61*(2), 87–105. https://doi.org/10.1016/j.cogpsych.2010.03.001

Chester, D. S. (2017). The role of positive affect in aggression. *Current Directions in Psychological Science*, *26*(4), 366–370. https://doi.org/10.1177/0963721417700457

Chester, D. S., Clark, M. A., & Dewall, C. N. (2021). The flux, pulse, and spin of aggression-related affect. *Emotion*. *21*(3), 513–525. https://doi.org/10.1037/emo0000730

Chester, D. S., & DeWall, C. N. (2016). The pleasure of revenge: Retaliatory aggression arises from a neural imbalance toward reward. *Social Cognitive and Affective Neuroscience*, *11*(7), 1173–1182. https://doi.org/10.1093/scan/nsv082

Chester, D. S., & DeWall, C. N. (2017). Combating the sting of rejection with the pleasure of revenge: A new look at how emotion shapes aggression. *Journal of Personality and Social Psychology*, *112*(3), 413–430. https://doi.org/10.1037/PSPI0000080

Chester, D. S., DeWall, C. N., & Enjaian, B. (2019). Sadism and Aggressive Behavior: Inflicting Pain to Feel Pleasure. *Personality and Social Psychology Bulletin*, *45*(8), 17. https://doi.org/doi.org/10.1177/0146167218816327

Cikara, M. (2015). Intergroup Schadenfreude: Motivating participation in collective violence. *Current Opinion in Behavioral Sciences*, *3*, 12–17. https://doi.org/10.1016/j.cobeha.2014.12.007

Cikara, M., Bruneau, E. G., & Saxe, R. R. (2011). Us and them: Intergroup failures of empathy. *Current Directions in Psychological Science*, *20*(3), 149–153. https://doi.org/10.1177/0963721411408713

Cikara, M., Bruneau, E., Van Bavel, J. J., & Saxe, R. (2014). Their pain gives us pleasure: How intergroup dynamics shape empathic failures and counter-empathic responses. *Journal of Experimental Social Psychology*, *55*, 110–125. https://doi.org/10.1016/j.jesp.2014.06.007

Cone, J., Brown-Iannuzzi, J. L., Lei, R., & Dotsch, R. (2020). Type I Error Is Inflated in the Two-Phase Reverse Correlation Procedure. *Social Psychological and Personality Science*, 194855062093861. https://doi.org/10.1177/1948550620938616

Darley, J. M., & Pittman, T. S. (2003). The Psychology of Compensatory and Retributive Justice. *Personality and Social Psychology Review*, *7*(4), 324–336. https://doi.org/10.1207/S15327957PSPR0704_05

Davis, F. C., Johnstone, T., Mazzulla, E. C., Oler, J. A., & Whalen, P. J. (2010). Regional response differences across the human amygdaloid complex during social conditioning. *Cerebral Cortex*, *20*(3), 612–621. https://doi.org/10.1093/cercor/bhp126

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*(1), 113–126. https://doi.org/10.1037/0022-3514.44.1.113.

de Quervain, D. J. F. (2004). The neural basis of altruistic punishment. *Science*, *305*(5688), 1254–1258. https://doi.org/10.1126/science.1100735

Decety, J., Chen, C., Harenski, C., & Kiehl, K. (2013). An fMRI study of affective perspective taking in individuals with psychopathy: Imagining another in pain does not evoke empathy. *Frontiers in Human Neuroscience*, *7*. https://www.frontiersin.org/article/10.3389/fnhum.2013.00489

Decety, J., Chen, C.-Y., & Cheng, Y. (2008). Gender differences in the mu rhythm during empathy for pain: An electroencephalographic study. *Brain Research,* 1251, 176-184. https://doi.org/10.1016/j.brainres.2008.11.062

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

DeWall, C. N., Anderson, C. A., & Bushman, B. J. (2012). Aggression. Chapter in I. Weiner (Ed.), *Handbook of Psychology*, 2nd Edition, Volume 5, 449-466. H. Tennen & J. Suls (Eds.), Personality and Social Psychology. New York: Wiley.

Dong, Y., Liu, Y., Jia, Y., Li, Y., & Li, C. (2018). Effects of facial expression and facial gender on judgment of trustworthiness: The modulating effect of cooperative and competitive settings. *Frontiers in Psychology*, *9*, 2022. https://doi.org/10.3389/fpsyg.2018.02022

Dotsch, R., & Todorov, A. (2012). Reverse Correlating Social Face Perception. *Social Psychological and Personality Science*, *3*(5), 562–571. https://doi.org/10.1177/1948550611430272

Dotsch, R., Wigboldus, D. H. J., Langner, O., & Van Knippenberg, A. (2008). Ethnic out-group faces are biased in the prejudiced mind. *Psychological Science*, *19*(10), 978–980. https://doi.org/10.1111/j.1467-9280.2008.02186.x

Drescher, A., & Schultheiss, O. C. (2016). Meta-analytic evidence for higher implicit affiliation and intimacy motivation scores in women, compared to men. *Journal of Research in Personality*, *64*, 1–10. https://doi.org/10.1016/j.jrp.2016.06.019

Drimalla, H., Landwehr, N., Hess, U., & Dziobek, I. (2019). From face to face: The contribution of facial mimicry to cognitive and emotional empathy. *Cognition and Emotion*, *33*(8), 1672–1686. https://doi.org/10.1080/02699931.2019.1596068

Duchaine, B., & Yovel, G. (2015). A Revised Neural Framework for Face Processing. *Annual Review of Vision Science*, *1*(1), 393–416. https://doi.org/10.1146/annurev-vision-082114-035518

Eadeh, F. R., Peak, S. A., & Lambert, A. J. (2017). The bittersweet taste of revenge: On the negative and positive consequences of retaliation. *Journal of Experimental Social Psychology*, *68*, 27–39. https://doi.org/10.1016/j.jesp.2016.04.007

Eder, A. B., Krishna, A., & Mitschke, V. (2021). Sweet Revenge Feels Less Bitter: Spontaneous Affective Reactions After Revenge Taking. *Social Psychological and Personality Science*, *12*(7), 1184–1192. https://doi.org/10.1177/19485506211019923

Eder, A. B., Mitschke, V., & Gollwitzer, M. (2020). What stops revenge taking? Effects of observed emotional reactions on revenge seeking. *Aggressive Behavior*, *46*(4), 305–316. https://doi.org/10.1002/ab.21890

Einarsen, S. (2000). Harassment and bullying at work. *Aggression and Violent Behavior*, *5*(4), 379–401. https://doi.org/10.1016/S1359-1789(98)00043-3

Eisenberg, N. (2000). Emotion, Regulation, and Moral Development. *Annual Review of Psychology*, *51*(1), 665–697. https://doi.org/10.1146/annurev.psych.51.1.665

Eisenberg, N., & Miller, P. A. (1987). The relation of empathy to prosocial and related behaviors. *Psychological Bulletin*, *101*(1), 91–119. https://doi.org/10.1037/0033-2909.101.1.91

Eitam, B., Kennedy, P. M., & Tory Higgins, E. (2013). Motivation from control. *Experimental Brain Research*, *229*(3), 475–484. https://doi.org/10.1007/s00221-012-3370-7

Ekman, P., & Friesen, W. V. (1982). Felt, false, and miserable smiles. *Journal of Nonverbal Behavior*, *6*(4), 238–252. https://doi.org/10.1007/BF00987191

Elson, M., Mohseni, M., 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

English Standard Version Bible (2001). Crossway. https://esv.literalword.com/

Falvello, V., Vinson, M., Ferrari, C., & Todorov, A. (2015). The Robustness of Learning about the Trustworthiness of Other People. *Social Cognition*, *33*(5), 368–386. https://doi.org/10.1521/soco.2015.33.5.368

Fanning, J. R., Keedy, S., Berman, M. E., Lee, R., & Coccaro, E. F. (2017). Neural Correlates of Aggressive Behavior in Real Time: A Review of fMRI Studies of Laboratory Reactive Aggression. *Current Behavioral Neuroscience*

Reports, *4*(2), 138–150. https://doi.org/10.1007/s40473-017-0115-8

Ferguson, C. J., & Rueda, S. M. (2009). Examining the validity of the modified Taylor competitive reaction time test of aggression. *Journal of Experimental Criminology*, *5*(2), 121–137. https://doi.org/10.1007/s11292-009-9069-5

Ferguson, C. J., Smith, S., Miller-Stratton, H., Fritz, S., & Heinrich, E. (2008). Aggression in the Laboratory: Problems with the Validity of the Modified Taylor Competitive Reaction Time Test as a Measure of Aggression in Media Violence Studies. *Journal of Aggression, Maltreatment & Trauma*, *17*(901740598), 118–132. https://doi.org/10.1080/10926770802250678

Fischer, A. H., Becker, D., & Veenstra, L. (2012). Emotional mimicry in social context: The case of disgust and pride. *Frontiers in Psychology*, *3*(NOV), 1–9. https://doi.org/10.3389/fpsyg.2012.00475

Fischer, A. H., Kret, M. E., & Broekens, J. (2018). Gender differences in emotion perception and self-reported emotional intelligence: A test of the emotion sensitivity hypothesis. *PLoS One*, *13*(1), e0190712. https://doi.org/10.1371/journal.pone.0190712

Forgas, J. P., & Bower, G. H. (1987). Mood Effects on Person-Perception Judgments. *Journal of Personality and Social Psychology*, *53*(1), 53–60. https://doi.org/10.1037/0022-3514.53.1.53

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review*, *118*(2), 247–279. https://doi.org/10.1037/a0022327

Fridlund, A. J., & Cacioppo, J. T. (1986). Guidelines for human electromyographic research. *Psychophysiology*, *23*(5), 567–589. https://doi.org/10.1111/j.1469-8986.1986.tb00676.x

Frijda, N. H. (1994). The lex talionis: On vengeance. In S. H. M. van Goozen, N. E. Van, de Poll & J. A. Sergeant (Eds.), *Emotions: Essays on emotion theory* (pp. 263–289). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Fujimura, T., Sato, W., & Suzuki, N. (2010). Facial expression arousal level modulates facial mimicry. *International Journal of Psychophysiology*, *76*(2), 88–92. https://doi.org/10.1016/j.ijpsycho.2010.02.008

Funk, F., McGeer, V., & Gollwitzer, M. (2014). Get the message: Punishment is satisfying if the transgressor responds to its communicative intent. *Personality and Social Psychology Bulletin*, *40*(8), 986–997. https://doi.org/10.1177/0146167214533130

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*(5), 692–731. https://doi.org/10.1037/0033-2909.132.5.692

Geniole, S. N., Keyes, A. E., Mondloch, C. J., Carré, J. M., & McCormick, C. M. (2012). Facing aggression: Cues differ for female versus male faces. *PLoS One*, *7*(1), 1–12. https://doi.org/10.1371/journal.pone.0030366

Gilbert, D. T., Pinel, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. P. (1998). Immune neglect: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology*, *75*(3), 617–638. https://doi.org/10.1037/0022-3514.75.3.617

Gollwitzer, M., & Bushman, B. J. (2012). Do Victims of Injustice Punish to Improve Their Mood? *Social Psychological and Personality Science*, *3*(5), 572–580. https://doi.org/10.1177/1948550611430552

Gollwitzer, M., & Denzler, M. (2009). What makes revenge sweet: Seeing the offender suffer or delivering a message? *Journal of Experimental Social Psychology*, *45*(4), 840–844. https://doi.org/10.1016/j.jesp.2009.03.001

Gollwitzer, M., Meder, M., & Schmitt, M. (2011). What gives victims satisfaction when they seek revenge? *European*

Journal of Social Psychology, *41*(3), 364–374. https://doi.org/10.1002/ejsp.782

Goubert, L., Craig, K. D., Vervoort, T., Morley, S., Sullivan, M. J. L., Williams, A. C. D. C., Cano, A., & Crombez, G. (2005). Facing others in pain: The effects of empathy. *Pain*, *118*(3), 285–288. https://doi.org/10.1016/j.pain.2005.10.025

Gross, J. J., & Feldman Barrett, L. (2011). Emotion Generation and Emotion Regulation: One or Two Depends on Your Point of View. *Emotion Review*, *3*(1), 8–16. https://doi.org/10.1177/1754073910380974

Han, S., Fan, Y., & Mao, L. (2008). Gender difference in empathy for pain: an electrophysiological investigation. *Brain research, 1196, 85-93.* https://doi.org/10.1016/j.brainres.2007.12.062

Hareli, S., & Hess, U. (2010). What emotional reactions can tell us about the nature of others: An appraisal perspective on person perception. *Cognition and Emotion*, *24*(1), 128–140. https://doi.org/10.1080/02699930802613828

Hareli, S., & Weiner, B. (2002). Dislike and envy as antecedents of pleasure at another's misfortune. *Motivation and Emotion*, *26*(4), 257–277. https://doi.org/10.1023/A:1022818803399

Haxby, J., Hoffman, E., & Gobbini, M. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–233. https://doi.org/10.1016/S1364-6613(00)01482-0

Healey, M. L., & Grossman, M. (2018). Cognitive and Affective Perspective-Taking: Evidence for Shared and Dissociable Anatomical Substrates. *Frontiers in Neurology*, *9*. https://doi.org/10.3389/fneur.2018.00491

Hess, U. (2021). Who to whom and why: The social nature of emotional mimicry. *Psychophysiology*, *58*(1), e13675. https://doi.org/10.1111/psyp.13675

Hess, U., Adams, R. B., & Kleck, R. E. (2005). Who may frown and who should smile? Dominance, affiliation, and the display of happiness and anger. *Cognition and Emotion*, *19*(4), 515–536. https://doi.org/10.1080/02699930441000364

Hess, U., Arslan, R., Mauersberger, H., Blaison, C., Dufner, M., Denissen, J. J. A., & Ziegler, M. (2017). Reliability of surface facial electromyography. *Psychophysiology*, *54*(1), 12–23. https://doi.org/10.1111/psyp.12676

Hess, U., & Hareli, S. (2017). The social signal value of emotions: The role of contextual factors in social inferences drawn from emotion displays. In J.-M. Fernández-Dols & J. A. Russell (Eds.), *The science of facial expression* (pp. 375–393). Oxford University Press.

Heyes, C. (2001). Causes and consequences of imitation. *Trends in Cognitive Sciences*, *5*(6), 253–261. https://doi.org/10.1016/S1364-6613(00)01661-2

Holland, A. C., O'Connell, G., & Dziobek, I. (2021). Facial mimicry, empathy, and emotion recognition: A meta-analysis of correlations. *Cognition and Emotion*, *35*(1), 150–168. https://doi.org/10.1080/02699931.2020.1815655

Horstmann, G. (2003). What do facial expressions convey: Feeling states, behavioral intentions, or action requests? *Emotion*, *3*(2), 150–166. https://doi.org/10.1037/1528-3542.3.2.150

Jackson, J. C., Choi, V. K., & Gelfand, M. J. (2019). Revenge: A multilevel review and synthesis. *Annual Review of Psychology*, *70*(1), 319–345. https://doi.org/10.1146/annurev-psych-010418-103305

Jäncke, L. (1994). An EMG investigation of the coactivation of facial muscles during the presentation of affect-laden stimuli. *Journal of Psychophysiology*, *8*(1), 1–10.

Kaiser, C. R., Vick, S. B., & Major, B. (2004). A Prospective Investigation of the Relationship Between Just-World Beliefs and the Desire for Revenge After September 11,

2001. *Psychological Science*, *15*(7), 503–506. https://doi.org/10.1111/j.0956-7976.2004.00709.x

Karremans, J. C., Dotsch, R., & Corneille, O. (2011). Romantic relationship status biases memory of faces of attractive opposite-sex others: Evidence from a reverse-correlation paradigm. *Cognition*, *121*(3), 422–426. https://doi.org/10.1016/j.cognition.2011.07.008

Kip, A., Blom, D., & van der Weiden, A. (2021). On the course of goal pursuit: The influence of goal progress on explicit judgments of self-agency. *Consciousness and Cognition*, *96*, 103222. https://doi.org/10.1016/j.concog.2021.103222

Kirkegaard Thomsen, D. (2006). The association between rumination and negative affect: A review. *Cognition and Emotion*, *20*(8), 1216–1235. https://doi.org/10.1080/02699930500473533

Knutson, B. (1996). Facial expressions of emotion influence interpersonal trait inferences. *Journal of Nonverbal Behavior*, *20*(3), 165–182. https://doi.org/10.1007/BF02281954

Koo, M., & Fishbach, A. (2014). Dynamics of self-regulation: How (un)accomplished goal actions affect motivation. *Motivation Science*, *1*(S), 73–90. https://doi.org/10.1037/2333-8113.1.S.73

Korb, S., Grandjean, D., & Scherer, K. R. (2010). Timing and voluntary suppression of facial mimicry to smiling faces in a Go/NoGo task—An EMG study. *Biological Psychology*, *85*(2), 347–349. https://doi.org/10.1016/j.biopsycho.2010.07.012

Korb, S., Malsert, J., Rochas, V., Rihs, T. A., Rieger, S. W., Schwab, S., Niedenthal, P. M., & Grandjean, D. (2015). Gender differences in the neural network of facial mimicry of smiles – An rTMS study. *Cortex*, *70*, 101–114. https://doi.org/10.1016/j.cortex.2015.06.025

Krämer, U. M., Büttner, S., Roth, G., & Münte, T. F. (2008). Trait aggressiveness modulates neurophysiological correlates of laboratory-induced reactive aggression in humans. *Journal of cognitive neuroscience*, *20*(8), 1464–1477. https://doi.org/10.1162/jocn.2008.20103

Krämer, U. M., Jansma, H., Tempelmann, C., & Münte, T. F. (2007). Tit-for-tat: the neural basis of reactive aggression. *Neuroimage*, *38*(1), 203-211. https://doi.org/10.1016/j.neuroimage.2007.07.029

Krumhuber, E., Manstead, A. S. R., Cosker, D., Marshall, D., Rosin, P. L., & Kappas, A. (2007). Facial Dynamics as Indicators of Trustworthiness and Cooperative Behavior. *Emotion*, *7*(4), 730–735. https://doi.org/10.1037/1528-3542.7.4.730

Kumar, D., & Srinivasan, N. (2014). Naturalizing Sense of Agency with a Hierarchical Event-Control Approach. *PLoS One*, *9*(3), e92431. https://doi.org/10.1371/journal.pone.0092431

Lambert, A. J., Peak, S. A., Eadeh, F. R., & Schott, J. P. (2014). How do you feel now? On the perceptual distortion of extremely recent changes in anger. *Journal of Experimental Social Psychology*, *52*, 82–95.

Lange, J., & Boecker, L. (2019). Schadenfreude as Social-Functional Dominance Regulator. *Emotion*, *19*(3), 489–502. https://doi.org/10.1037/emo0000454.supp

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the radboud faces database. *Cognition and Emotion*, *24*(8), 1377–1388. https://doi.org/10.1080/02699930903485076

Lanzetta, J. T., & Englis, B. G. (1989). Expectations of Cooperation and Competition and Their Effects on Observers' Vicarious Emotional Responses. *Journal of Personality and Social Psychology*, *56*(4), 543–554. https://doi.org/10.1037/0022-3514.56.4.543

Larsen, J. T., Norris, C. J., & Cacioppo, J. T. (2003). Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology*, *40*(5), 776–785. https://doi.org/10.1111/1469-8986.00078

Larsen, R. J. (2000). Toward a Science of Mood Regulation. *Psychological Inquiry*, *11*(3), 129–141. https://doi.org/10.1207/S15327965PLI1103_01

Leach, C. W., Spears, R., Branscombe, N. R., & Doosje, B. (2003). Malicious Pleasure: Schadenfreude at the Suffering of Another Group. *Journal of Personality and Social Psychology*, *84*(5), 932–943. https://doi.org/10.1037/0022-3514.84.5.932

Lefevre, C. E., & Lewis, G. J. (2014). Perceiving Aggression from Facial Structure: Further Evidence for a Positive Association with Facial Width-to-Height Ratio and Masculinity, but not for Moderation by Self-Reported Dominance. *European Journal of Personality*, *28*(6), 530–537. https://doi.org/10.1002/per.1942

Lepron, E., Causse, M., & Farrer, C. (2014). Responsibility and the sense of agency enhance empathy for pain. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1799). https://doi.org/10.1098/rspb.2014.2288

Likowski, K. U., Mühlberger, A., Seibt, B., Pauli, P., & Weyers, P. (2008). Modulation of facial mimicry by attitudes. *Journal of Experimental Social Psychology*, *44*(4), 1065–1072. https://doi.org/10.1016/j.jesp.2007.10.007

Likowski, K. U., Weyers, P., Seibt, B., Stöhr, C., Pauli, P., & Mühlberger, A. (2011). Sad and lonely? Sad mood suppresses facial mimicry. *Journal of Nonverbal Behavior*, *35*(2), 101–117. https://doi.org/10.1007/s10919-011-0107-4

Lotze, M., Veit, R., Anders, S., & Birbaumer, N. (2007). Evidence for a different role of the ventral and dorsal medial prefrontal cortex for social reactive aggression: An interactive fMRI study. *NeuroImage*, *34*(1), 470–478. https://doi.org/10.1016/j.neuroimage.2006.09.028

Markovic, J., Anderson, A. K., & Todd, R. M. (2014). Tuning to the significant: Neural and genetic processes underlying affective enhancement of visual perception and memory. *Behavioural Brain Research*, *259*, 229–241. https://doi.org/10.1016/j.bbr.2013.11.018

Martin, J., Rychlowska, M., Wood, A., & Niedenthal, P. (2017). Smiles as Multipurpose Social Signals. *Trends in Cognitive Sciences*, *21*(11), 864–877. https://doi.org/10.1016/j.tics.2017.08.007

Mattarozzi, K., Colonnello, V., Russo, P. M., & Todorov, A. (2019). Person information facilitates memory for face identity. *Psychological Research*, *83*(8), 1817–1824. https://doi.org/10.1007/s00426-018-1037-0

Mauersberger, H., & Hess, U. (2019). When smiling back helps and scowling back hurts: Individual differences in emotional mimicry are associated with self-reported interaction quality during conflict interactions. *Motivation and Emotion*, *43*(3), 471–482. https://doi.org/10.1007/s11031-018-9743-x

Mazur, A., & Booth, A. (1998). Testosterone and social dominance in men. *Behavioral and Brain Sciences*, *21*(3), 353–397. https://doi.org/DOI: undefined

Mehta, P. H., & Beer, J. (2010). Neural Mechanisms of the Testosterone–Aggression Relation: The Role of Orbitofrontal Cortex. *Journal of Cognitive Neuroscience*, *22*(10), 2357–2368. https://doi.org/10.1162/jocn.2009.21389

Mellentin, A. I., Dervisevic, A., Stenager, E., Pilegaard, M., & Kirk, U. (2015). Seeing enemies? A systematic review of anger bias in the perception of facial expressions among anger-prone and aggressive populations. *Aggression and Violent Behavior*, *25*, 373–383. https://doi.org/10.1016/j.avb.2015.09.001

Mende-Siedlecki, P. (2018). Changing our minds: The neural bases of dynamic impression updating. In *Current Opinion in Psychology* (Bd. 24, S. 72–76). Elsevier B.V. https://doi.org/10.1016/j.copsyc.2018.08.007

Merlhiot, G., Mondillon, L., Méot, A., Dutheil, F., & Mermillod, M. (2021). Facial width-to-height ratio underlies perceived dominance on facial emotional expressions. *Personality and Individual Differences*, *172*, 110583. https://doi.org/10.1016/j.paid.2020.110583

Miller, D. T. (2001). Disrespect and the Experience of Injustice. *Annual Review of Psychology*, *52*(1), 527–553. https://doi.org/10.1146/annurev.psych.52.1.527

Mitschke, V., & Eder, A. B. (2021). Facing the enemy: Spontaneous facial reactions towards suffering opponents. *Psychophysiology*, *58*(8). https://doi.org/10.1111/psyp.13835

Moody, E. J., McIntosh, D. N., Mann, L. J., & Weisser, K. R. (2007). More than mere mimicry? The influence of emotion on rapid facial reactions to faces. *Emotion*, *7*(2), 447–457. https://doi.org/10.1037/1528-3542.7.2.447

Murata, A., Saito, H., Schug, J., Ogawa, K., & Kameda, T. (2016). Spontaneous facial mimicry is enhanced by the goal of inferring emotional states: Evidence for moderation of „automatic“ mimicry by higher cognitive processes. *PLoS One*, *11*(4). https://doi.org/10.1371/journal.pone.0153128

Nasby, W., Hayden, B., & DePaulo, B. M. (1980). Attributional bias among aggressive boys to interpret unambiguous social stimuli as displays of hostility. *Journal of Abnormal Psychology*, *89*(3), 459–468. https://doi.org/10.1037/0021-843X.89.3.459

Nayak, S., Shiflett, S. C., Eshun, S., & Levine, F. M. (2000). Culture and gender effects in pain beliefs and the prediction of pain tolerance. *Cross-Cultural Research*, *34*(2), 135–151. https://doi.org/10.1177/106939710003400203

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

Niedenthal, P. M., Mermillod, M., Maringer, M., & Hess, U. (2010). The Simulation of Smiles (SIMS) model: Embodied simulation and the meaning of facial expression. *Behavioral and Brain Sciences*, *33*(6), 417–433. https://doi.org/10.1017/S0140525X10000865

Oliveira, M., Garcia-Marques, T., Dotsch, R., & Garcia-Marques, L. (2019). Dominance and competence face to face: Dissociations obtained with a reverse correlation approach. *European Journal of Social Psychology*, *49*(5), 888–902. https://doi.org/10.1002/ejsp.2569

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(32), 11087–11092. https://doi.org/10.1073/pnas.0805664105

Osgood, J. M. (2017). Is revenge about retributive justice, deterring harm, or both? *Social and Personality Psychology Compass*, *11*(1), e12296. https://doi.org/10.1111/spc3.12296

Paulus, C. (2009). *Der Saarbrücker Persönlichkeitsfragebogen SPF (IRI) zur Messung von Empathie: Psychometrische Evaluation der deutschen Version des interpersonal reactivity index.* http://psydok.sulb.uni-saarland.de/volltexte/2009/2363/

Peterson, C. K., Gravens, L. C., & Harmon-Jones, E. (2011). Asymmetric frontal cortical activity and negative affective responses to ostracism. *Social Cognitive and Affective Neuroscience*, *6*(3), 277–285. https://doi.org/10.1093/scan/nsq027

Plant, E. A., Hyde, J. S., Keltner, D., & Devine, P. G. (2000). The Gender Stereotyping of Emotions. *Psychology of Women Quarterly*, *24*(1), 81–92. https://doi.org/10.1111/j.1471-6402.2000.tb01024.x

Pillutla, M. M., & Murnighan, J. K. (1996). Unfairness, Anger, and Spite: Emotional Rejections of Ultimatum Offers. *Organizational Behavior and Human Decision Processes*, 68(3), 208–224. https://doi.org/10.1006/obhd.1996.0100

Platt, T. (2008). Emotional responses to ridicule and teasing: Should gelotophobes react differently? *Humor - International Journal of Humor Research*, *21*(2). https://doi.org/10.1515/HUMOR.2008.005

Ramírez, J. M., & Andreu, J. M. (2006). Aggression, and some related psychological constructs (anger, hostility, and impulsivity) Some comments from a research project. *Neuroscience and Biobehavioral Reviews*, *30*(3), 276–291. https://doi.org/10.1016/j.neubiorev.2005.04.015

Ratner, K. G., Dotsch, R., Wigboldus, D. H. J., Van Knippenberg, A., & Amodio, D. M. (2014). Visualizing minimal ingroup and outgroup faces: implications for impressions, attitudes, and behavior. *Journal of personality and social psychology, 106(6),* 897. https://doi.org/10.1037/a0036498

Ren, D., Wesselmann, E. D., & Williams, K. D. (2018). Hurt people hurt people: Ostracism and aggression. *Current Opinion in Psychology*, *19*(17), 34–38. https://doi.org/10.1016/j.copsyc.2017.03.026

Repple, J., Pawliczek, C. M., Voss, B., Siegel, S., Schneider, F., Kohn, N., & Habel, U. (2017). From provocation to aggression: The neural network. *BMC Neuroscience 2017 18:1*, *18*(1), 1–9. https://doi.org/10.1186/S12868-017-0390-Z

Roberton, T., Daffern, M., & Bucks, R. S. (2012). Emotion regulation and aggression. *Aggression and Violent Behavior*, *17*(1), 72–82. https://doi.org/10.1016/j.avb.2011.09.006

Rule, N. O., Slepian, M. L., & Ambady, N. (2012). A memory advantage for untrustworthy faces. *Cognition*, *125*(2), 207–218. https://doi.org/10.1016/j.cognition.2012.06.017

Rychlowska, M., Jack, R. E., Garrod, O. G. B., Schyns, P. G., Martin, J. D., & Niedenthal, P. M. (2017). Functional Smiles: Tools for Love, Sympathy, and War. *Psychological Science*, *28*(9), 1259–1270. https://doi.org/10.1177/0956797617706082

Rychlowska, M., Manstead, A. S. R., & Schalk, J. van der. (2019). The Many Faces of Smiles. *The Social Nature of Emotion Expression*, 227–245. https://doi.org/10.1007/978-3-030-32968-6_13

Sachisthal, M. S. M., Sauter, D. A., & Fischer, A. H. (2016). Mimicry of ingroup and outgroup emotional expressions. *Comprehensive Results in Social Psychology*, *1*(1–3), 86–105. https://doi.org/10.1080/23743603.2017.1298355

Sato, W., Fujimura, T., & Suzuki, N. (2008). Enhanced facial EMG activity in response to dynamic facial expressions. *International Journal of Psychophysiology*, *70*(1), 70–74. https://doi.org/10.1016/j.ijpsycho.2008.06.001

Schmid, P. C., Mast, M. S., Bombari, D., Mast, F. W., & Lobmaier, J. S. (2011). How mood states affect information processing during facial emotion recognition: An eye tracking study. *Swiss Journal of Psychology*, *70*(4), 223–231. https://doi.org/10.1024/1421-0185/a000060

Seibt, B., Mühlberger, A., Likowski, K. U., & Weyers, P. (2015). Facial mimicry in its social setting. *Frontiers in Psychology*, *6*(August). https://doi.org/10.3389/fpsyg.2015.01122

Sessa, P., & Meconi, F. (2015). Perceived trustworthiness shapes neural empathic responses toward others' pain. *Neuropsychologia*, *79*, 97–105. https://doi.org/10.1016/j.neuropsychologia.2015.10.028

Shah, R. I. A. B. R., & Tee, E. Y. (2019). Identifying with the In-group Increases Aggressive Tendencies against the Outgroup: The Mediating Role of Schadenfreude. In *Emotions and Leadership*. Emerald Publishing Limited. https://doi.org/10.1108/s1746-979120190000015011

Shergill, S. S., Bays, P. M., Frith, C. D., & Wolpert, D. M. (2003). Two Eyes for an Eye: The Neuroscience of Force Escalation. *Science*, *301*, 187–187. https://doi.org/10.1126/science.1085327

Sidarus, N., Vuorre, M., & Haggard, P. (2017). Integrating prospective and retrospective cues to the sense of agency: A multi-study investigation†. *Neuroscience of Consciousness*, *3*(1), 1–10. https://doi.org/10.1093/nc/nix012

Singer, T., & Klimecki, O. M. (2014). Empathy and compassion. *Current Biology*, *24*(18), R875–R878. https://doi.org/10.1016/j.cub.2014.06.054

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for Pain Involves the Affective but not Sensory Components of Pain. *Science*, *303*(5661), 1157–1162. https://doi.org/10.1126/science.1093535

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature*, *439*(7075), 466–469. https://doi.org/10.1038/nature04271

Smeijers, D., Rinck, M., Bulten, E., van den Heuvel, T., & Verkes, R. J. (2017). Generalized hostile interpretation bias regarding facial expressions: Characteristic of pathological aggressive behavior. *Aggressive Behavior*, *43*(4), 386–397. https://doi.org/10.1002/ab.21697

Smith, R. H., & van Dijk, W. W. (2018). Schadenfreude and Gluckschmerz. *Emotion Review*, *10*(4), 293–304. https://doi.org/10.1177/1754073918765657

Sonnby-Borgström, M., Jönsson, P., & Svensson, O. (2003). Emotional empathy as related to mimicry reactions at different levels of information processing. *Journal of Nonverbal Behavior*, *27*(1), 3–23. https://doi.org/10.1023/A:1023608506243

Stillwell, A. M., Baumeister, R. F., & Del Priore, R. E. (2008). We're All Victims Here: Toward a Psychology of Revenge. *Basic and Applied Social Psychology*, *30*(3), 253–263. https://doi.org/10.1080/01973530802375094

Strelan, P. (2018). Justice and Forgiveness in Interpersonal Relationships. *Current Directions in Psychological Science*, *27*(1), 20–24. https://doi.org/10.1177/0963721417734311

Sun, Y.-B., Wang, Y.-Z., Wang, J.-Y., & Luo, F. (2015). Emotional mimicry signals pain empathy as evidenced by facial electromyography. *Scientific Reports*, *5*(1), 16988. https://doi.org/10.1038/srep16988

Sutherland, C. A. M., Oldmeadow, J. A., & Young, A. W. (2016). Integrating social and facial models of person perception: Converging and diverging dimensions. *Cognition*, *157*, 257–267. https://doi.org/10.1016/j.cognition.2016.09.006

Szameitat, D. P., Alter, K., Szameitat, A. J., Darwin, C. J., Wildgruber, D., Dietrich, S., & Sterr, A. (2009). Differentiation of emotions in laughter at the behavioral level. *Emotion*, *9*(3), 397–405. https://doi.org/10.1037/a0015692

Tamir, M. (2021). Effortful Emotion Regulation as a Unique Form of Cybernetic Control. *Perspectives on Psychological Science*, *16*(1), 94–117. https://doi.org/10.1177/1745691620922199

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

Tice, D. M., & Bratslavsky, E. (2000). Giving in to Feel Good: The Place of Emotion Regulation in the Context of

General Self-Control. *Psychological Inquiry*, *11*(3), 149–159. https://doi.org/10.1207/S15327965PLI1103_03

Tiedens, L. Z. (2001). Anger and advancement versus sadness and subjugation: The effect of negative emotion expressions on social status conferral. *Journal of Personality and Social Psychology*, *80*(1), 86–94. https://doi.org/10.1037/0022-3514.80.1.86

Tripp, T. M., Bies, R. J., & Aquino, K. (2007). A Vigilante Model of Justice: Revenge, Reconciliation, Forgiveness, and Avoidance. *Social Justice Research*, *20*(1), 10–34. https://doi.org/10.1007/s11211-007-0030-3

Trope, Y. (1986). Identification and Inferential Processes in Dispositional Attribution. *Psychological Review*, *93*(3), 239–257. https://doi.org/10.1037/0033-295X.93.3.239

Vachon, D. D., Lynam, D. R., & Johnson, J. A. (2014). The (Non)relation between empathy and aggression: Surprising results from a meta-analysis. *Psychological Bulletin*, *140*(3), 751–773. https://doi.org/10.1037/a0035236

van der Schalk, J., Fischer, A., Doosje, B., Wigboldus, D., Hawk, S., Rotteveel, M., & Hess, U. (2011). Convergent and divergent responses to emotional displays of ingroup and outgroup. *Emotion*, *11*(2), 286–298. https://doi.org/10.1037/A0022582

van 't Wout, M., & Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition*, *108*(3), 796–803. https://doi.org/10.1016/j.cognition.2008.07.002

Vanman, E. J. (2016). The role of empathy in intergroup relations. *Current Opinion in Psychology*, *11*, 59–63. https://doi.org/10.1016/j.copsyc.2016.06.007

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.

West, S. J., Lasko, E. N., Hall, C. J., Khan, N. G., & Chester, D. S. (2022). Some revenge now or more revenge later? Applying an intertemporal framework to retaliatory aggression. Motivation Science, 8(1), 33-35. https://doi.org/10.1037/mot0000248

Weyers, P., Mühlberger, A., Kund, A., Hess, U., & Pauli, P. (2009). Modulation of facial reactions to avatar emotional faces by nonconscious competition priming. *Psychophysiology*, *46*(2), 328–335. https://doi.org/10.1111/j.1469-8986.2008.00771.x

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

Wilson, J. P., & Rule, N. O. (2015). Facial Trustworthiness Predicts Extreme Criminal-Sentencing Outcomes. *Psychological Science*, *26*(8), 1325–1331. https://doi.org/10.1177/0956797615590992

Wise, E. A., Price, D. D., Myers, C. D., Heft, M. W., & Robinson, M. E. (2002). Gender role expectations of pain: Relationship to experimental pain perception. *Pain*, *96*(3), 335–342. https://doi.org/10.1016/S0304-3959(01)00473-0

Wood, A., Rychlowska, M., Korb, S., & Niedenthal, P. (2016). Fashioning the Face: Sensorimotor Simulation Contributes to Facial Expression Recognition. *Trends in Cognitive Sciences*, *20*(3), 227–240. https://doi.org/10.1016/j.tics.2015.12.010

Wyer, R. S., Jr. (2007). Principles of mental representation. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (pp. 285–307). The Guilford Press.

Yoshimura, S. (2007). Goals and emotional outcomes of revenge activities in interpersonal relationships. *Journal of Social and Personal Relationships*, *24*(1), 87–98. https://doi.org/10.1177/0265407507072592

Young, K. A., Gandevia, S. C., & Giummarra, M. J. (2017). Vicarious pain responders and emotion: Evidence for distress rather than mimicry. *Psychophysiology*, *54*(7), 1081–1095. https://doi.org/10.1111/psyp.12865

Zaki, J. (2014). Empathy: A motivated account. *Psychological Bulletin*, *140*(6), 1608–1647. https://doi.org/10.1037/a0037679

Appendix A

Supplement Chapter 2

Fixed Effect Omnibus Tables Study 1

Fixed Effect Omnibus tests Zygomaticus Major

	F	Num df	Den df	p
Provocation	27.664	1	1059	< .001
Emotion	12.785	1	1059	< .001
Time	0.871	6	1059	0.515
Provocation ∗ Emotion	16.527	1	1059	< .001
Provocation ∗ Time	0.221	6	1059	0.970
Emotion ∗ Time	0.486	6	1059	0.819

Fixed Effect Omnibus tests Corrugator Supercilii

	F	Num df	Den df	p
Provocation	6.945	1	1059	0.009
Emotion	0.269	1	1059	0.604
Time	0.223	6	1059	0.970
Provocation ∗ Emotion	5.692	1	1059	0.017
Provocation ∗ Time	0.602	6	1059	0.729
Emotion ∗ Time	0.421	6	1059	0.866

Fixed Effect Omnibus Tables Study 2

Fixed Effect Omnibus tests Zygomaticus Major

	F	Num df	Den df	p
Provocation	41.200	1	2933	< .001
Emotion	3.313	2	2933	0.037
Time	4.330	11	2933	< .001
Provocation ✻ Emotion	13.604	2	2933	< .001
Provocation ✻ Time	2.011	11	2933	0.024
Emotion ✻ Time	0.417	22	2933	0.992

Fixed Effect Omnibus tests Orbiculars Oculi

	F	Num df	Den df	p
Provocation	14.96	1	2933	< .001
Emotion	12.22	2	2933	< .001
Time	13.44	11	2933	< .001
Provocation ✻ Emotion	5.20	2	2933	0.006
Provocation ✻ Time	2.46	11	2933	0.005
Emotion ✻ Time	1.53	22	2933	0.054

Fixed Effect Omnibus tests Corrugator Supercilii

	F	Num df	Den df	p
Provocation	2.013	1	2933	0.156
Emotion	6.315	2	2933	0.002
Time	1.398	11	2933	0.166
Provocation ✻ Emotion	37.313	2	2933	< .001
Provocation ✻ Time	0.338	11	2933	0.977
Emotion ✻ Time	0.523	22	2933	0.966

Female Subset Analysis Study 1

Study 1 Fixed Effect Omnibus Tests with female subset only (N= 35).

Fixed Effect Omnibus tests Zygomaticus Major

	F	Num df	Den df	p
Provocation	**33.172**	**1**	**924**	**< .001**
Emotion	**5.649**	**1**	**924**	**0.018**
Time	0.988	6	924	0.432
Provocation * Emotion	**18.807**	**1**	**924**	**< .001**
Provocation * Time	0.221	6	924	0.970
Emotion * Time	0.214	6	924	0.973

Fixed Effect Omnibus tests Corrugator Supercilii

	F	Num df	Den df	p
Provocation	**6.833**	**1**	**924**	**0.009**
Emotion	0.327	1	924	0.568
Time	0.238	6	924	0.964
Provocation * Emotion	**5.801**	**1**	**924**	**0.016**
Provocation * Time	0.555	6	924	0.766
Emotion * Time	0.403	6	924	0.878

Female Subset Analysis Study 2

Study 2 Fixed Effect Omnibus Tests with female subset only (N= 35).

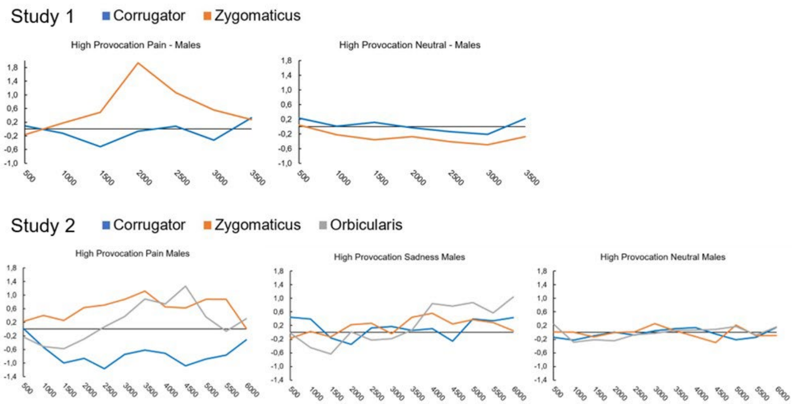
Fixed Effect Omnibus tests **Zygomaticus Major**	F	Num df	Den df	*p*
Provocation	26.245	1	2436	< .001
Emotion	3.810	2	2436	0.022
Time	3.976	11	2436	< .001
Provocation ∗ Emotion	12.755	2	2436	< .001
Provocation ∗ Time	1.792	11	2436	0.050
Emotion ∗ Time	0.408	22	2436	0.993

Fixed Effect Omnibus tests **Orbicularis Oculi**	F	Num df	Den df	*p*
Provocation	7.990	1	2436.000	0.005
Emotion	7.340	2	2436.000	< .001
Time	11.225	11	2436.000	< .001
Provocation ∗ Emotion	6.592	2	2436.000	0.001
Provocation ∗ Time	1.699	11	2436.000	0.068
Emotion ∗ Time	1.129	22	2436.000	0.306

Fixed Effect Omnibus tests **Corrugator Supercilii**	F	Num df	Den df	*p*
Provocation	0.022	1	2436.000	0.883
Emotion	**3.880**	**2**	**2436.000**	**0.021**
Time	1.527	11	2436.000	0.115
Provocation ∗ Emotion	**23.529**	**2**	**2436.000**	**< .001**
Provocation ∗ Time	0.279	11	2436.000	0.990
Emotion ∗ Time	0.513	22	2436.000	0.970

Figure S1

Male Subset Figure



Time course of the z-transformed fEMG response of the male subset ($N_1 = 5$; $N_2 = 7$).

Dispositional Measures of Empathy

Study 1

Correlations of ZM activity after full pain expression onset (1000ms-3000ms) after provocation with the empathic concern score were significant (r = -0.439, *p* = 0.005). Tests of the sub facets were conducted using Bonferroni adjusted alpha levels of .0125 per test (.05/4). Please see Supplement Table 1 and 2 for other sub facets. This negative correlation signified that higher positive signal differences (stronger ZM activations) were associated with lower scores on empathic concern. No significant correlation was observed for the empathic concern score and CS activity. The Cronbach's α values in Study 1 were .54 for empathic concern, .77 for fantasy, .79 for personal distress and .64 for perspective taking.

Further, a correlation analysis between the empathy sub scales and noise level (punishment) resulted in significant negative correlations between empathic concern and noise level unprovoked, r = -.339, p = .034, as well as between perspective taking and noise level unprovoked, r = -.321, p = .046 (see Supplement Table 3 for all sub facets and inter correlations). None of the dispositional empathy sub facets was related to the choice of noise level during provocation blocks.

Table S1

Mean SPF scores study 1 and 2

	Study 1	Study 2
Empathic Concern	15.74 (2.07)	14.95 (2.43)
Perspective Taking	14.38 (3.13)	14.02 (3.48)
Fantasy	13.87 (3.59)	13.43 (3.61)
Personal Distress	10.85 (3.18)	10.48 (3.10)

Table S2

Correlation between ZM activity following pain displays and empathy sub facets in Study 1

Variables	1	2	3	4	5	6
1. Mean ZM activation	–					
2. Empathic Concern	-.404*	–				
3. Perspective Taking	-.238	.333*	–			
4. Personal Distress	-.091	-.149	-.084	–		
5. Fantasy	.112	0.301	-.078	.218	–	
6. Total IRI	-.242	.612**	.490**	.463**	.682**	–

*$p < .05$, **$p < .01$, ***$p < .001$.

Table S3

Correlations between SPF sub facets and punishment noise level in Study 1

Variables	1	2	3	4	5	6	7
1. Empathic Concern	–						
2. Perspective Taking	.333*	–					
3. Personal Distress	-.149	-.084	–				
4. Fantasy	0.301	-.078	.218	–			
5. Total IRI	.612**	.490**	.463**	.682**	–		
6. Noise Level Provoked	-.224	-.179	.116	-.021	-.123	–	
7. Noise Level Unprovoked	-.339*	-.321*	.175	-.131	-.258	.748**	–

*p < .05, **p < .01, ***p < .001.

Study 2

To investigate the possible influence of dispositional empathy on punishment behavior, SPF subscale scores were correlated with noise level, which yielded a negative correlation between empathic concern and level of noise punishment during provocation, $r = -0.334$, $p = .031$ (see Supplement Table 4 for all sub facets and inter correlations).). The correlation did not withstand the Bonferroni adjusted alpha level of .0125. The Cronbach's α values in Study 2 were .43 for empathic concern, .73 for fantasy, .67 for personal distress and .86 for perspective taking (for a descriptive summary of the empathy subscales see Table 1 in the Supplement).

Table S4

Correlations between SPF sub facets and punishment noise level in Study 2

Variables	1	2	3	4	5	6	7
1. Empathic Concern	–						
2. Perspective Taking	.323*	–					
3. Personal Distress	.139	-.228	–				
4. Fantasy	.317*	.331*	.069	–			
5. Total IRI	.669**	.621**	.377**	.750**	–		
6. Noise Level Provoked	-.334*	-.275	-.024	.090	-.198	–	
7. Noise Level Unprovoked	-.381*	-.084	-.135	-.031	-.228	.526**	–

*p < .05, **p < .01, ***p < .001.

Pre-registered comparisons between displays of pain and sadness

Direct comparisons of muscle activity following pain and sadness displays yielded no significant differences. Comparisons of CS activity between pain and sadness reactions revealed no significant effect of *Emotion*, $F(1,41) = 1.60$, $p = .213$, neither did the comparison of ZM activity, $F(1,41) = 2.83$, $p = .100$ and OO activity, $F(1,41) = 2.38$, $p = .131$.

Video Materials

All video materials were chosen based on their scores on the dimensions pain and sadness. The video clips can be downloaded at https://osf.io/ysnd3/. Ratings were obtained by an independent sample of 289 participants, divided in 3 subsets. Each subset originally contained 60 videos. Each videos was rated on each emotion dimension (“Please rate how intensely the person in the video might feel [emotion]”) on unipolar scales ranging from 1=“not at all” to 5=“extremely”. For a more detailed description please see Eder et al. (2020).

Table S5

Emotional ratings of the opponent expressions of sadness selected for Study 1 and 2

Video Type	Model no.	Video no.	Sadness *M (SD)*	Pain *M (SD)*
Sadness Expression	1	1	3.46(1.21)	2.46 (1.17)
		2	3.38 (1.32)	2.94 (1.31)
	5	1	2.78 (1.10)	1.86 (1.05)
		2	2.70 (1.15)	2.00 (1.10)
	18	1	3.51 (1.14)	2.15 (1.21)
		2	2.77 (1.10)	1.69 (0.85)
	20	1	3.60 (1.13)	2.10 (1.10)
		2	3.63 (1.17)	2.30 (1.17)
	21	1	3.23 (1.14)	2.08 (1.14)
		2	3.26 (1.33)	1.69 (0.90)
	26	1	2.99 (1.26)	1.93 (1.19)
		2	2.71 (1.20)	2.01 (1.20)
	overall		3,17 (0.36)	2.10 (0.35)

Table S6

Emotional ratings of the opponent expressions of pain selected for Study 1 and 2

Video Type	Model no.	Video no.	Sadness *M (SD)*	Pain *M (SD)*
Pain Expression	1	1	2.39 (1.14)	3.00 (1.07)
		2	1.78 (1.00)	3.10 (1.39)
	5	1	2.19 (1.06)	3.35 (1.10)
		2	2.12 (1.01)	2.88 (1.26)
	6	1	2.41 (1.26)	3.61 (1.36)
		2	2.50 (1.31)	3.04 (1.27)
	12	1	1.46 (0.79)	2.28 (1.11)
		2	2.28 (1.03)	2.55 (1.10)
	18	1	1.84 (0.96)	2.45 (1.24)
		2	1.75 (1.00)	3.01 (1.25)
	20	1	2.07 (1.11)	3.30 (1.28)
		2	2.50 (1.16)	3.71 (1.20)
	21	1	1.83 (0.90)	2.53 (1.15)
		2	1.81 (1.10)	2.76 (1.21)
	26	1	1.60 (0.85)	2.39 (1.04)
		2	1.89 (1.00)	2.82 (1.21)
	overall		2.03 (0.33)	2.92 (0.43)

Appendix B

Supplement Chapter 4

Comparison of AntiCIs and CIs obtained in Study 1

Provocation. Comparison of provocation Anti-Classification Images to Classification Images created in Study 1, resulted in significant differences in trustworthiness, $t(132) = 7.04$, $p < .001$, $dz = 0.61$, with the AntiCI being rating as more trustworthy ($M = 3.41$, $SD = 1.45$), than the provocation CI ($M = 2.37$, SD = 1.22), as well as more likeable, $t(132) = 10.62$, $p < .001$, $dz = 0.92$. A comparison of masculinity, $t(132) = 0.31$, $p = .758$, $dz = 0.03$ and dominance, $t(132) = 0.94$, $p = 0.349$, $dz = 0.08$, did not result in any significant differences between the AntiCi and CI. Please see Table S1 for an overview of descriptive data.

No Provocation. Comparison of no provocation Anti-Classification Images to Classification Images created in Study 1, resulted in significant differences in trustworthiness, $t(132) = 5.61$, $p < .001$, $dz = 0.49$, with the AntiCI being rating as more trustworthy ($M = 3.69$, $SD = 1.39$), than the provocation CI ($M = 2.79$, SD = 1.34), as well as more likeable, $t(132) = 7.50$, $p < .001$, $dz = 0.65$. A comparison of masculinity, $t(132) = 0.31$, $p = .758$, $dz = 0.03$ and dominance, $t(132) = 1.68$, $p = .096$, $dz = 0.15$, did not result in any significant differences between the AntiCi and CI.

Table S1

Anti-Classification Image Rating Means (*SDs*) in Study 1

CI Type	trustworthiness	likability	dominance	masculinity
Provocation	3.41 (1.45)	3.68 (1.40)	3.71 (1.38)	4.14 (1.40)
No provocation	3.69 (1.39)	3.89 (1.40)	3.50 (1.45)	4.11 (1.38)

Table S2

Correlations between rating dimensions Study 1.

Variables	1.	2.	3.	4.
1. Trustworthiness	–			
2. Likability	0.511***	–		
3. Dominance	0.322**	0.365***	–	
4. Masculinity	0.404***	0.406***	0.745***	–

Note. * $p < .05$. ** $p < .01$. *** $p < .001$

Table S3

Correlations between rating dimensions Study 2.

Variables	1.	2.	3.	4.
1. Trustworthiness	–			
2. Hostility	0.701***	–		
3. Dominance	0.524***	0.550***	–	
4. Valence	0.579***	0.581***	0.482***	–

Note. * $p < .05$. ** $p < .01$. *** $p < .001$

Exploratory Analysis of Gender Effects

Mixed ANOVAs with the within factor *CI Type* (provocation, no provocation) and the between factor *Gender* revealed significant differences in masculinity and dominance ratings based on rater gender. The mixed ANOVA of masculinity ratings yielded a significant main effect of *Gender*, $F(1,131) = 6.16$, $p = 0.014$, $\eta_p^2 = .045$. Post hoc tests indicate that female participants rated CIs as less masculine than male raters did (mean difference = -0.69, $p_{tukey} = .014$, $t = -2.48$), irrespective of condition. A mixed ANOVA of dominance ratings yielded a significant interaction between *CI Type* and *Gender*, $F(1,131) = 6.13$, $p = 0.015$, $\eta_p^2 = .045$. While female raters show no difference in dominance ratings, male raters assigned higher levels of dominance to the CI of the unprovoked condition. Ratings of trustworthiness and likability were not influenced by rater gender, all Fs $\leqslant 0.26$, all ps $\geqslant .68$.

Table S4

Mean ratings (with SD) of aggressive motives in the provocation blocks divided by condition.

	I wanted to impair my opponent's performance in order to win more.	I wanted to control my opponent's level or responses.	I wanted to drive my opponent mad.	I wanted to hurt my opponent.	I wanted to pay back my opponent for the noise levels he set.	I wanted to blast him harder than he blasted me.
1 pain	1.43(0.84)	2.40(1.03)	1.78(1.23)	1.85(1.23)	1.58(1.03)	2.68(1.53)
2 anger	1.78(1.03)	2.68(1.10)	2.08(1.27)	2.05(1.28)	1.51(0.72)	2.85(1.15)

Note. Motives were rated after each game round on unipolar scales from 1 = not at all to 5 = a lot.

Table S5

SPF Empathy scores divided by condition (1 = pain, 2 = anger).

		Mean (*SD*)	Min-Max
Empathic Concern	1	15.05(3.09)	7-20
	2	16.17(2.13)	11-20
Perspective Taking	1	15.13(2.86)	10-20
	2	14.88(2.79)	8-20
Fantasy	1	12.13(4.30)	5-20
	2	15.08(3.48)	7-20
Personal Distress	1	12.05(2.76)	7-18
	2	11.78(3.10)	5-18
Total score	1	56.35(8.48)	40-75
	2	57.90(7.28)	41-74

Table S6

Anger and Sympathy ratings divided by condition (1 = pain, 2 = anger)

		Mean (*SD*)
Anger towards opponent	1	3.25(1.39)
	2	3.15(1.17)
Sympathy	1	2.48(1.01)
	2	2.20(0.91)
Justice Satisfaction	1	8.43(2.50)
	2	8.53(2.80)

Note. Attitudes were rated on unipolar scales from 1 = not at all to 5 = a lot. Justice Satisfaction is based on an index of 3 items, with a maximum possible score of 15.

Exploratory Analysis of Forced Choice Comparison Study 2

As an exploratory additional analysis for Study 2, we asked the rating sample to directly compare both classification images (CI) obtained from each condition. CI images were shown next to each other, and participants were asked which of the two images was more trustworthy (negative, hostile, dominant). Table S7 presents the results from this direct comparison task.

Table S7

Binary choice comparisons via binomial tests.

	Level	Count	Total	Proportion	p
dominant	Pain	140	292	0.479	0.520
	Anger	152	292	0.521	0.520
hostile	Pain	108	292	0.370	< .001
	Anger	184	292	0.630	< .001
negative	Pain	112	292	0.384	< .001
	Anger	180	292	0.616	< .001
trustworthy	Pain	162	292	0.555	0.069
	Anger	130	292	0.445	0.069

Note. H_a is proportion ≠ 0.5

Appendix C

Anteile der Koautoren kumulativer Teil der Dissertation

Mitschke, V. & Eder, A.B. (2021). Facing the enemy: Spontaneous facial reactions towards suffering opponents. *Psychophysiology*, 58(8). https://doi.org/10.1111/psyp.13835.

Vanessa Mitschke: Konzeptualisierung, Vorbereitung und Durchführung der Experimente; Analyse und Interpretation der Daten; Schreiben des originalen und revidierten Manuskripts; Korrespondenz während des Review- und Produktionsprozesses

Datum Unterschrift

Andreas Eder: Konzeptualisierung der Experimente; Bereitstellung finanzieller und technischer Ressourcen; Feedback zu originalem und revidiertem Manuskript

Datum Unterschrift

Appendix D

Lebenslauf

Aus Datenschutzgründen wurde der Lebenslauf entfernt.