AKADEMIA PEDAGOGIKI SPECJALNEJ
im. Marii Grzegorzewskiej

Marta Bodecka-Zych

Rozprawa doktorska na temat:

Poznawcze, osobowościowe i relacyjne czynniki związane z wrogimi atrybucjami

Rozprawa doktorska przygotowana pod kierunkiem

dr hab. Anny Zajenkowskiej, prof. APS

Warszawa 2022

Marta Bodecka-Zych

OŚWIADCZENIE

I. Świadom(a) odpowiedzialności prawnej oświadczam, że złożona rozprawa doktorska na temat „Poznawcze, osobowościowe i relacyjne czynniki związane z wrogimi atrybucjami" została przygotowana przeze mnie przy wykorzystaniu wykazanej w tej rozprawie literatury przedmiotu i materiałów źródłowych.

II. Oświadczam jednocześnie, że rozprawa ta nie narusza praw autorskich, dóbr osobistych chronionych prawem cywilnym oraz nie zawiera informacji i danych uzyskanych w sposób niedozwolony prawem i nie była dotychczas przedmiotem żadnej procedury, związanej z uzyskaniem dyplomów lub tytułów zawodowych uczelni wyższej.

III. Oświadczam ponadto, że niniejsza wersja rozprawy doktorskiej jest identyczna z załączoną wersją elektroniczną.

IV. Wyrażam zgodę na udostępnianie egzemplarza mojej rozprawy doktorskiej w czytelni Biblioteki Głównej Akademii Pedagogiki Specjalnej im. Marii Grzegorzewskiej w Warszawie.

Data …….. ……..

Oświadczenie promotora rozprawy:

Niniejsza rozprawa jest gotowa do oceny przez recenzentów.

Data 30.11.2022 ……..

Streszczenie rozprawy doktorskiej

Imiona i nazwisko autora rozprawy: Marta Bodecka-Zych
Stopień /tytuł naukowy oraz imiona i nazwisko promotora rozprawy:
dr hab. Anna Zajenkowska, prof. APS
Temat rozprawy doktorskiej:
Poznawcze, osobowościowe i relacyjne czynniki związane z wrogimi atrybucjami
Słowa kluczowe: wrogie atrybucje, przetwarzanie informacji społecznych, ogólny model agresji, wrażliwość na prowokację , narcyzm wrażliwy, osoby osadzone

Streszczenie

Celem projektu było pogłębienie i uzupełnienie wiedzy na temat poznawczych, osobowościowych i relacyjnych korelatów wrogich atrybucji. Jest to temat ważny naukowo i społecznie, ponieważ dokonywanie wrogich atrybucji, czyli interpretowanie zachowań innych osób jako intencjonalnych i wrogich, jest jednym z najważniejszych czynników wyjaśniających różnice indywidualne w częstości i nasileniu reaktywnych zachowań agresywnych. Jednakże dyskusja na temat tego, z jakimi zmiennymi osobowymi łączą się wrogie atrybucje, na ile relacje te są stabilne w zależności od kontekstu społeczno-relacyjnego oraz co prowadzi do wykształcenia się wrogiego stylu atrybucyjnego, pozostaje cały czas otwarta.

Niniejsza rozprawa doktorska opiera się na cyklu siedmiu artykułów opublikowanych w recenzowanych czasopismach naukowych. Dzięki zastosowaniu zróżnicowanych metod (m.in. pomiaru wrogich atrybucji) i technik badawczych (np. technika eyetrackingu) oraz uwzględnieniu różnych grup (np. kobiet i mężczyzn skazanych za przestępstwa z użyciem przemocy, par ojców i synów, pacjentów) prace te dostarczają nowych i istotnych informacji na temat wrogich atrybucji, stanowią także przyczynek do polemiki naukowej.

W artykułach 1 i 2, które powstały we współpracy z międzynarodowym zespołem badawczym, przyglądaliśmy się wzorcom kodowania i interpretowania niejednoznacznych sygnałów społecznych przez osoby skazane za przestępstwa z użyciem przemocy. Jako jedni z nielicznych przeprowadziliśmy tego typu badania nie tylko wśród osadzonych mężczyzn, ale także kobiet. Wyniki opisane w dwóch pierwszych artykułach pokazują między innymi, że osoby skłonne do agresji wykazują deficyty w zakresie uważności na oczy i twarze, co może skutkować trudnościami w adekwatnym rozumieniu sytuacji społecznych.

Artykuły 3 i 4 uwzględniają badania dotyczące osobowościowych korelatów wrogich atrybucji. W artykule 3 opisaliśmy różnice w zakresie cechy gniewu i wrażliwości na prowokację, a także historii uwięzienia członków rodziny osób badanych wśród kobiet i mężczyzn przebywających w zakładach karnych oraz osób z populacji ogólnej. Z kolei w artykule 4 poszukiwaliśmy odpowiedzi na pytanie, czy wrażliwość na prowokację przewiduje tendencję do interpretowania niejednoznacznych ekspresji emocjonalnych twarzy jako rozgniewanych.

Artykuły 5 i 6 traktują o relacyjnych czynnikach związanych z wrogimi atrybucjami. W ramach prowadzonych badań zweryfikowaliśmy związek narcyzmu wrażliwego z podwyższonym poziomem wrogich atrybucji w zależności od kontekstu społeczno-relacyjnego oraz sprawdziliśmy, czy poziom wrogich atrybucji ojca koreluje z poziomem wrogich atrybucji jego syna oraz czy znaczenie dla tej relacji ma poziom narcyzmu wrażliwego ojca. Z kolei w artykule 7 znajduje się opis metody pozwalającej na redukcję tendencji do dokonywania wrogich atrybucji.

Wyniki zrealizowanych badań ujęte w cyklu 7 artykułów uzupełniają model dotyczący wrogich atrybucji o kilka znaczących elementów, podkreślają znaczenie m.in. płci, kontekstu społeczno-relacyjnego czy poziomu ambiwalencji ocenianej sytuacji.

Potwierdzają także przypuszczenie, że dokonywanie wrogich atrybucji jest procesem złożonym i zasadne jest prowadzenie kolejnych badań, które pozwolą na pełniejsze zrozumienie tego zjawiska i projektowanie skutecznych działań praktycznych mających na celu ich redukcję.

Summary of the Ph.D. Thesis

Name of the author of the thesis: Marta Bodecka-Zych
Academic title of the doctoral supervisor of the thesis:
Dr hab. Anna Zajenkowska, prof. APS
The title of the thesis:
Cognitive, individual, and relational factors related to hostile attributions
Key words phrases: Hostile attributions, social information processing, sensitivity to provocation, anger trait, vulnerable narcissism, inmates

Abstract

The aim of the current project was to improve and complement our knowledge of the cognitive, individual, and relational correlates of hostile attributions, a scientifically and socially important topic because making hostile attributions – that is, interpreting the behaviour of others as intentional and hostile – is one of the most important factors explaining individual differences in the frequency and severity of reactive-aggressive behaviour. However, the questions of which particular, personal variables are associated with hostile attributions, how stable these relationships are (depending on their socio-relational context), and what leads to the development of a hostile attributional style, all remain open.

This dissertation is based on a series of seven articles, published in peer-reviewed scientific journals, employing a diverse set of methodologies (e.g., different methods of measuring hostile attributions) and research techniques (e.g., eye tracking technique). Study participants included a variety of socio-demographic groups (e.g., men and women convicted of violent crimes, father-son pairs, and patients). Collectively, an analysis of these papers provides new and relevant information to the scholarly discussion of hostile attributions, particularly those that constitute an element of scientific polemics.

In papers 1 and 2, written in collaboration with an international research team, we examine patterns of encoding and interpreting ambiguous social signals, by people convicted of violent crimes. Here, we conduct one of the few research studies of this type, not only among men but also among female inmates. The results described in the first two articles demonstrate, among other things, that those prone to aggression show deficits in attentiveness to eyes and faces, which can result in difficulty in adequately understanding social situations.

Papers 3 and 4 represent a line of research on the personality correlates of hostile attributions. In article 3, we describe differences in terms of trait anger and sensitivity to provocation, as well as the history of incarceration in the families of the subjects studied. Subsequently, in paper 4, we examine whether sensitivity to provocation predicts the tendency to interpret ambiguous emotional, facial expressions as angry.

In papers 5 and 6, we deal with the relational factors associated with hostile attributions. We verify the association of vulnerable narcissism with elevated levels of hostile attributions, according to the socio-relational context, and examine whether a father's level of hostile attributions correlates with his son's level of hostile attributions. We also examine the relevance of a father's level of vulnerable narcissism to this relationship. In turn, in paper 7, we describe a method for reducing the tendency to make hostile attributions.

The results of the research add several significant elements to the model on hostile attributions, emphasizing the importance of gender, socio-relational context, and the level of the ambiguity in the assessed situation, among other factors. The findings also confirm the assumption that making hostile attributions is a very complex process, and further research is warranted to more fully understand the phenomenon—and to enable the design of effective, practical measures to reduce hostile attributions.

Spis treści

Podziękowania

Nie sposób podziękować wszystkim osobom, które towarzyszyły mi podczas nauki w szkole doktorskiej i przyczyniły się do powstania tej pracy. Jednakże chciałabym wymienić tych, bez których przejście tej drogi byłoby dużo trudniejsze i z pewnością mniej owocne.

Składam najserdeczniejsze podziękowania mojej promotorce – dr hab. Annie Zajenkowskiej. prof. APS. Aniu, przez lata naszej współpracy byłaś wspaniałą przewodniczką po naukowym świecie, zawsze wspierającą, ale przy tym wymagająca wymagającą i dodającą odwagi. Dziękuję, że tak wiele mogłam się od Ciebie nauczyć.

Dziękuje osobom tworzącym szkołę doktorską, Pani kierownik dr hab. Aleksandrze Tłuściak-Deliowskiej, prof. APS. Radzie Naukowej, moim kolegom i koleżankom. Szczególne podziękowania kieruje do osób, z którymi rozpoczynałam ten etap kształcenia. Dagmaro, Iwono, Krzysztofie, Adrianno, Karolu, Edyto i Adrianie, cieszę się, że wspólnie przecieraliśmy szlaki.

Chciałabym podziękować także całemu zespołowi Zakładu Psychologii Społecznej, jestem bardzo wdzięczna, że mogłam się od Was uczyć jako studentka i doktorantka. Dziękuję, że zawsze mogę liczyć na Wasze wsparcie.

I w końcu, dziękuję mojej rodzinie i przyjaciołom, że towarzyszyli mi w tej naukowej drodze, a przede wszystkim mojemu mężowi, który najlepiej wie, jak kręta bywała ta droga.

1. Wykaz artykułów

Niniejsza rozprawa doktorska podejmuje tematykę wrogich atrybucji. Opiera się na cyklu siedmiu artykułów opublikowanych w recenzowanych, międzynarodowych i polskich czasopismach naukowych. Kierowałam powstawaniem czterech artykułów, w których jestem wiodącą autorką, w przypadku pozostałych tekstów jestem jedną z osób, które przyczyniły się do ich przygotowania i opublikowania. Cykl ten składa się z następujących prac:

Artykuł 1. Gehrer, N.A., Zajenkowska, A., Bodecka, M., & Schönenberg, M. (2021). Attention orienting to the eyes in violent female and male offenders: An eye-tracking study. *Biological Psychology, 163,* 108136. IF 3.111; 100 pkt MEiN.

Artykuł 2. Zajenkowska, A. M., Bodecka, M., Duda, E., & Lawrence, C. (2022). Reduced attention toward faces, intentionality and blame ascription in violent offenders and community based adults: Evidence from an eye-tracking study. *Aggressive Behavior. 48*(2), 264-275. IF 3.047; 100 pkt MEiN.

Artykuł 3. Bodecka-Zych, M., Zajenkowska, A., & Bower Russa, M. (2022). Sex Differences in Inmates: Anger, Sensitivity to Provocation and Family History of Imprisonment. *International Journal of Offender Therapy and Comparative Criminology, 66*(12), 1327-1342. IF 1.645; 70 pkt MEiN.

Artykuł 4. Zajenkowska, A., Bodecka-Zych, M., Gehrer, N., Krejtz, K., Lawrence, C., Schönenberg, M., & Jusyte, A. (2022). Gender differences in sensitivity to provocation and hostile attribution bias toward ambiguous facial cues in violent offenders and community-based adults. *Motivation and Emotion*, 1–10. IF 4.135; 100 pkt MEiN.

Artykuł 5. Bodecka-Zych, M., Jonason, P. K., & Zajenkowska, A. (2022). Hostile attribution biases in vulnerable narcissists depends on the socio-relational context. *Journal of Individual Differences, 43*(2), 70–78. IF 2.608; 70 pkt MEiN.

Artykuł 6. Bodecka-Zych, M., Zajenkowska, A., & Lawrence, C. (2022). Dad, are they laughing at me? Fathers' vulnerable narcissism and sons' hostile attributions. *Personality and Individual Differences, 192*, 111582. IF 3.950; 100 pkt MEiN.

Artykuł 7. Bodecka, M., Jakubowska, A., Zajenkowska, A. (2021). Warsztat psychoedukacyjny z elementami mentalizacji i jego rola w redukowaniu wrogości wśród osób osadzonych. [Psycho-educational workshop with elements of mentalization and its role in reducing hostility among inmates] *Psychologia Wychowawcza, 62*(20), 101–115. 40 pkt MEiN.

2. Wstęp

Różnice indywidualne w zachowaniach agresywnych są funkcją charakterystycznych stylów przypisywania (lub nie) wrogich intencji innym ludziom

(Dodge, 2006, s. 792)

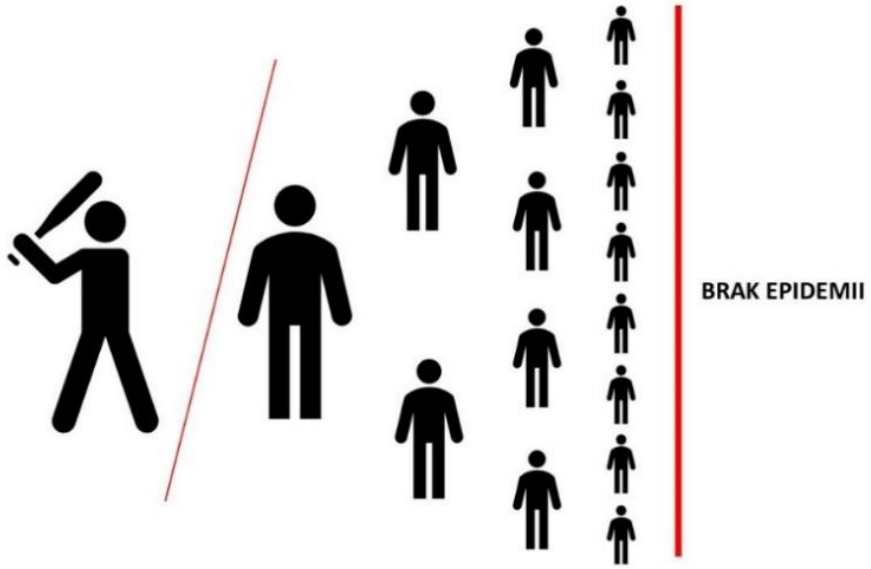
Ponad czterdzieści lat temu, Kenneth Dodge (1980), w artykule stworzonym na podstawie swojej pracy doktorskiej, dowiódł, że agresywni chłopcy w porównaniu z tymi nieagresywnymi są bardziej skłonni do przypisywania wrogich intencji rówieśnikom, nie ufają innym i szczególnie w niejednoznacznych sytuacjach – oczekują agresji z ich strony. W tym samym roku William Nasby i współpracownicy (1980) opublikowali artykuł dotyczący tendencji do przypisywania wrogości (*attributional bias to infer hostility*). Obydwie, wyżej wymienione prace uważane są za prekursorskie dla badań dotyczących dokonywania wrogich atrybucji, czyli **interpretowania zachowań innych osób jako intencjonalnych i wrogich, szczególnie w sytuacjach ambiwalentnych, czyli takich, w których trudno jednoznacznie ocenić intencje drugiego człowieka (Epps i Kendal, 1995).** Wrogie atrybucje są jednym z najważniejszych czynników wyjaśniających różnice indywidualne w częstości i nasileniu reaktywnych zachowań agresywnych (de Castro i in., 2002; Dodge, 2006; Martinneli i in., 2018; Tuente i in., 2019). Dzięki licznym badaniom znamy mechanizm łączący poznanie społeczne z zachowaniem (Crick i Dodge, 1994), a także potencjalne zmienne związane z dokonywaniem wrogich atrybucji. Im bogatszy jest jednak zbiór wiedzy na ten temat, tym więcej pytań się pojawia. Wcześniejsze badania wskazywały przede wszystkim na korelaty osobowościowe (np. cecha gniewu) i w zakresie wzorców przetwarzania poznawczego (np. deficyty w zakresie patrzenia na oczy). Mniej badań poświęcono zmiennym relacyjnym. Konieczne są więc dalsze badania dotyczące zarówno każdej

z wymienionych kategorii zmiennych, ale także badania odnoszące się do warunków, w jakich te związki się ujawniają.

Biorąc pod uwagę powyższe, celem mojej dotychczasowej pracy było: 1) zweryfikowanie poznawczych (np. wzorce kodowania wskazówek społecznych), osobowościowych (np. narcyzm wrażliwy, podatność na prowokację) oraz relacyjnych (relacja ojca z synem) korelatów wrogich atrybucji oraz 2) określenie roli innych zmiennych (np. kontekstu społeczno-relacyjnego, płci) w relacji pomiędzy tymi czynnikami a wrogimi atrybucjami.

3. Ramy teoretyczne projektu

Pandemia COVID-19 zdominowała rzeczywistość w ciągu ostatnich 3 lat. Liczne obostrzenia, ograniczenia kontaktów społecznych, stres związany z obawą o zdrowie i życie, wszystko to przełożyło się na jakość życia milionów ludzi. Naukowcy skupili się na zbadaniu wirusa, zrozumieniu mechanizmów jego działania i finalnie przygotowaniu szczepionki. W efekcie powstało prawie pięć milionów publikacji na temat COVID 19 (Google Scholar, 2022). Badacze koncentrują się w nich m.in. na kwestiach związanych z poczuciem zagrożenia życia, ale eksponują też fakt, że pandemia COVID-19 pojawiła się nagle i niespodziewanie. Istnieją jednak pandemie, które nie pojawiają się znienacka, przykładem takiego zjawiska jest przemoc, która rocznie dotyka milionów osób – zarówno dzieci, jak i dorosłych (rys.1). Podobnie jak w przypadku chorób zakaźnych, pandemia przemocy pojawia się w charakterystycznych skupiskach, wzrasta i opada w falowych konfiguracjach, przenosi się i rozprzestrzenia (Slutkin i in., 2015). Pandemia nie istnieje bez wirusa, czyli cząstki, która infekuje żywe organizmy (np. Dousari i in., 2020). W związku z tym, poszerzając wiedzę na temat przemocy, kluczowe jest zbadanie czynników, które do niej prowadzą.



Rysunek 1. Profilaktyka epidemii przemocy, opracowanie własne na podstawie materiałów prezentowanych przez G. Slutkina podczas konferencji ISRA 2022 w Kanadzie.

Najsilniejszym predyktorem przemocy jest poprzedni akt przemocy (Slutkin i in., 2015). Przy czym przemoc rozumiana jest jako powtarzające się zachowania agresywne, w których występuje nierównowaga sił pomiędzy sprawcą a ofiarą lub jako skrajna forma agresji, której celem jest wyrządzenie poważnej szkody fizycznej (Anderson i Bushman, 2002; Bushman i Huesmann, 2010; Huesmann i Taylor, 2006). Z kolei agresja w psychologii społecznej definiowana jest jako świadome i intencjonalne zachowanie, które ma na celu skrzywdzenie innej osoby, która stara się tej krzywdy uniknąć (Bushman i Huesmann, 2010; DeWall, i in., 2012). Każdy akt przemocy uważany jest za przypadek agresji, ale nie każdy akt agresji można nazywać przemocą (Allen i Anderson, 2017; Anderson i Bushman, 2002). Agresja jest więc konstruktem bardziej podstawowym. Aby poznać mechanizmy prowadzące do przemocy, konieczne jest poznanie przyczyn agresji.

3.1. Ogólny model agresji

Anderson i Bushman (2002) dwadzieścia lat temu połączyli główne, współcześnie istniejące teorie psychologiczne dotyczące ludzkiej agresywności w jeden ogólny model GAM (*General Aggression Model*). Model ten, osadzony w podejściu poznawczo-społecznym, uwzględnia zarówno czynniki indywidualne, jak i sytuacyjne związane z wyzwalaniem zachowań agresywnych. Do czynników sytuacyjnych autorzy zaliczyli m.in.: sytuacje prowokacji, frustracji, a także te sytuacje, które są związane z występowaniem bólu i dyskomfortu, czy też dostępnością agresywnych wskazówek, np. broni. Z kolei wśród czynników osobowych znalazły się wszelkie różnice indywidualne, mające znaczenie dla sposobu, w jaki jednostka reaguje na daną sytuacje – np. niestabilna samoocena, narcyzm, cecha gniewu, pozytywne nastawienie do agresji, niska samokontrola, a także tendencja do dokonywania wrogich atrybucji (Allen i in., 2018), czyli interpretowania zachowań innych osób jako intencjonalnych i wrogich (Epps i Kendal, 1995). Wrogie atrybucje zajmują ważne miejsce w modelu GAM (Allen i in., 2018). Dzięki wieloletnim badaniom wiemy, że odgrywają one szczególną rolę w przewidywaniu agresji reaktywnej (wrogiej, gorącej), która wiąże się z uczuciami gniewu i pojawia się w reakcji na postrzegane zagrożenie (Crick i Dodge, 1996). W związku z tym wrogie atrybucje są bezpośrednio związane ze sposobem, w jaki jednostka przetwarza informacje społeczne (Crick i Dodge, 1994).

3.2. Model przetwarzania informacji społecznych

Wrogi sposób interpretowania interakcji społecznych jest związany z kodowaniem wskazówek oraz znaczeniem, jakie jest im nadawane. Są to elementy teorii przetwarzania informacji społecznych (*Social Information Processing*; Crick i Dodge, 1994). Według tego modelu zachowanie jest funkcją sześciu, następujących po sobie

kroków (rys. 2), które mogą prowadzić do zachowania agresywnego. W swoim projekcie eksplorowałam zagadnienia związane z kodowaniem i interpretowaniem wskazówek społecznych, ponieważ to właśnie dwa początkowe kroki prowadzą do powstania mentalnej reprezentacji sytuacji społecznej (Crick i Dodge, 1994), kształtując tym samym dalsze przetwarzanie i efekt końcowy – zachowanie.

Rysunek 2. Model przetwarzania informacji społecznych (Crick i Dodge, 1994, s. 76), adaptacja na potrzeby artykułu: „Warsztat psychoedukacyjny z elementami mentalizacji i jego rola w redukowaniu wrogości wśród osób osadzonych".

Kodowanie wskazówek. Każda sytuacja społeczna składa się z dużej liczby potencjalnie istotnych informacji. Sprawne funkcjonowanie w przestrzeni społecznej jest warunkowane szybkim i selektywnym kodowaniem (Horsley i in., 2020). Ludzie różnią

się między sobą sposobem, w jaki wybierają i kodują informacje, ta sama sytuacja może być więc zakodowana w zupełnie inny sposób przez różne osoby (np. Crick i Dodge, 1994, Larkin i in., 2013). Istnieje wiele hipotez dotyczących wzorców kodowania związanych z reakcjami agresywnymi (m.in. Crick i Dodge, 1994; Horsley i in., 2010). Dwie główne i konkurencyjne zarazem dotyczyły pytania: co leży u podstaw kodowania? Czy jest to nadwrażliwość na wrogie i zagrażające wskazówki (hipoteza dół–góra)? Czy też nabyte wcześniej wrogie schematy przechowywane w pamięci (hipoteza góra–dół)? W celu udzielenia odpowiedzi na ww. pytania, prowadzono m.in. badania z wykorzystaniem techniki eyetrackingu (śledzenia ruchów gałek ocznych), która umożliwia stwierdzenie między innymi tego, na jakie elementy (np. ilustracji) patrzy osoba badana, które z nich przykuwają jej uwagę jako pierwsze, jak długo patrzy na poszczególne części obrazka. Okazało się, że agresywne dzieci skupiają wzrok dłużej na niewrogich wskazówkach, a mimo to przypisują więcej wrogości niż ich nieagresywni rówieśnicy (Horsley i in., 2010). Podobne rezultaty uzyskano w badaniach z udziałem osób dorosłych. Uczestnicy charakteryzujący się wysokim poziomem cechy gniewu dłużej skupiali wzrok na wskazówkach świadczących o nieumyślnym działaniu aktora, a mimo to oceniali jego działanie jako wrogie (Wilkowski i in., 2007). Zatem dokonywanie wrogich atrybucji wiąże się między innymi z szybką, automatyczną oceną sytuacji, opartą na wcześniej nabytych schematach, która może następować nawet przed zakodowaniem specyficznych (wrogich lub niewrogich) wskazówek dostępnych w danej sytuacji.

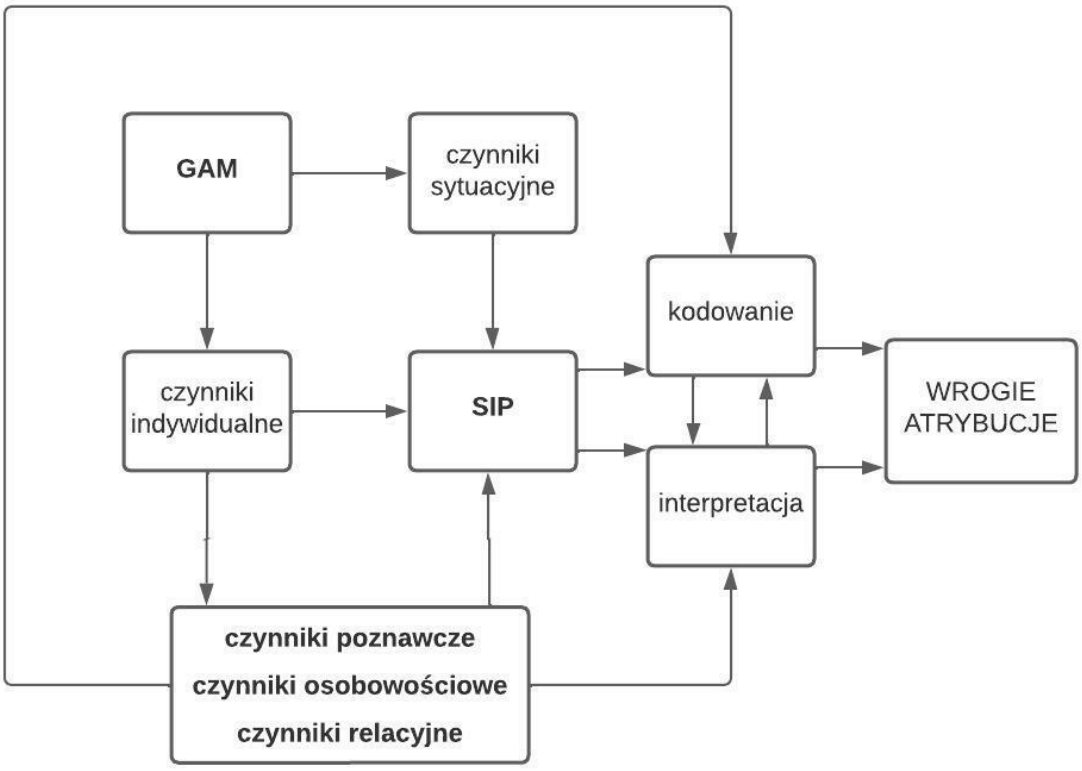
***Interpretowanie wskazówek*.** Krok ten jest kluczowy dla moich badań, ponieważ właśnie na tym etapie mogą pojawić się wrogie interpretacje niejednoznacznych sytuacji, przypisywanie nieprzyjaznych zamiarów innym ludziom, a także, jak pokazują ostatnie badania, obwinianie i gniewne uczucia wobec partnera/ partnerów interakcji (Combs i in., 2007; Crick i Dodge, 1994; 1996; Zajenkowska i in., 2021).

W większości opublikowanych prac wrogie atrybucje były traktowane jako jednolity konstrukt. Tymczasem istnieją badania (Combs i in., 2007; Zajenkowska i in., 2021), które wykazują, że możemy wskazać trzy subfaktory składające się na tę zmienną: pierwszym z nich jest przypisanie intencjonalności – zinterpretowanie zachowania drugiej osoby jako celowego i zamierzonego (np. osoba, z którą się umówiłeś(-łaś) nie przyszła na spotkanie, myślisz: „Zrobiła to celowo"). Drugim jest obwinianie drugiej osoby za negatywne konsekwencje zachowania („Gdyby chciała, to by przyszła, to jej wina, nic jej nie tłumaczy"), ostatnim elementem są pojawiające się gniewne uczucia („Jestem wściekły(-kła) na tę osobę, ja jej teraz pokażę"). Intencjonalność i przypisywanie winy, mimo że mogą pojawiać się w tym samym czasie, są odrębne w odniesieniu do gniewu. Wykazano na przykład, że w przypadku związku wrażliwości na prowokację z gniewem istotną rolę mediacyjną odgrywają wrogie atrybucje, jednakże samo przypisanie intencjonalności nie wystarczy, aby pojawił się gniew, konieczne jest obwinienie sprawcy (Zajenkowska i in., 2021). Wydaje się więc, że dokonywanie wrogich atrybucji jest procesem bardziej złożonym, niż do tej pory sądzono. Konieczne jest więc kontynuowanie badań, które pozwoliłyby na pełniejsze zrozumienie mechanizmów stojących za tym zjawiskiem.

4. Cel i podstawy teoretyczne badań własnych

Dyskusja naukowa na temat tego, jakie wzorce uwagowe i cechy osobowości łączą się z wrogimi atrybucjami, na ile relacje te są stabilne w zależności od kontekstu społeczno-relacyjnego oraz co łączy się z powstawaniem wrogiego stylu atrybucyjnego, jest cały czas otwarta. W związku z tym celem mojego projektu było zbadanie poznawczych, osobowościowych i relacyjnych czynników związanych z wrogimi atrybucjami. Swoje badania oparłam na ogólnym modelu agresji (GAM, Anderson i Bushman, 2002) oraz teorii przetwarzania informacji społecznych – SIP (Crick i Dodge,

1994). Podążając za autorami wyżej wymienionych teorii, założyłam, że zachowania agresywne są uwarunkowane strukturami wiedzy, które wpływają na wiele różnych zjawisk społeczno-poznawczych, w tym na percepcję i interpretację (Allen i in., 2018). Struktury wiedzy, w tym schematy poznawcze, powstają na skutek wcześniejszych doświadczeń jednostki – często związanych z ważnymi relacjami (Young, 1956; Dodge, 2006), a suma tych struktur składa się na jej osobowość (Allen i in., 2018). Dlatego też, wyjaśniając mechanizm powstawania i działania wrogich atrybucji (prowadzących do agresji), skupiam się na trzech rodzajach czynników (rys. 3) oraz warunkach, które mają znaczenie dla manifestowania się tych zmiennych w działaniu. Dodatkowo, na podstawie wyników dostępnych i zrealizowanych przeze mnie badań, poszukuję sposobów redukcji tendencji do dokonywania wrogich atrybucji.



Rysunek 3. Czynniki poznawcze, osobowościowe i relacyjne związane z wrogimi atrybucjami w kontekście modelu GAM i SIP, opracowanie własne.

4.1. Poznawcze korelaty wrogich atrybucji

Osoby mające problemy z agresywnymi reakcjami wykazują trudności w rozpoznawaniu emocji (Chapman i in., 2018; Marsh i Blair, 2008). Wykazano na przykład, że agresywni i antyspołeczni przestępcy mają trudności z rozpoznawaniem emocji strachu (Blair, 2001). Deficyt ten można częściowo tłumaczyć charakterystycznym wzorcem percepcyjnym – osoby agresywne zwracają mniejszą uwagę na oczy, które są ważnym nośnikiem informacji o stanie psychicznym drugiej osoby (Gehrer, 2018; 2020; Smith i in., 2005). Ta percepcyjna „niedokładność" może skutkować tym, że wiele sytuacji społecznych odbieranych jest jako niejednoznaczne, co z kolei aktywizuje wrogi schemat, prowadzi do większej sensytywności na oznaki zagrożenia (np. oznaki gniewu w ekspresji emocjonalnej twarzy) i w konsekwencji do dokonania wrogiej atrybucji (Schönenberg i Jusyte, 2014).

Jednakże badania dotyczące prawidłowości percepcyjnych osób charakteryzujących się agresywnością, dotychczas prowadzone były głównie wśród mężczyzn, niewiele wiadomo na temat tego, czy deficyty w zakresie patrzenia występują także w przypadku kobiet. Co więcej, w badaniach nad rozpoznawaniem afektu zwykle wykorzystywano bodźce przedstawiające wyłącznie twarze zwrócone przodem do osoby biorącej udział w badaniu, tymczasem ważne wydaje się zbadanie, w jaki sposób osoby agresywne odbierają bardziej złożone sytuacje, w których przedstawione są całe postacie – zarówno sprawcy działania, jak i odbiorcy (ofiary). Dlatego też w badaniach prowadzonych wraz z zespołem badawczym poszukiwałam odpowiedzi na pytanie, czy istnieją różnice międzypłciowe w zakresie koncentracji uwagi na oczach wśród kobiet i mężczyzn osadzonych za przestępstwa z użyciem przemocy **(artykuł 1)**. Ponadto zbadaliśmy, czy deficyty uwagowe związane z postrzeganiem twarzy pojawiają się także w przypadku, gdy prezentowane są całe postaci i czy rola, jaką osoba pełni w danej

sytuacji (osoba krzywdząca vs. krzywdzona), ma znaczenie dla wzorca percepcyjnego osoby obserwującej **(artykuł 2).** Dodatkowo zaproponowaliśmy nowy wskaźnik wrogich atrybucji – izomorfizm intencjonalności i winy.

4.2. Osobowościowe korelaty wrogich atrybucji

Wzorce percepcyjne mogą wiązać się z osobowością. Dotychczasowe wyniki badań wskazują na istnienie szeregu cech indywidualnych związanych z tendencją do dokonywania wrogich atrybucji. Do najbardziej znaczących należą cecha gniewu oraz wrażliwość na prowokację.

Gniew odgrywa rolę pośredniczącą pomiędzy wrogimi atrybucjami a agresywnym zachowaniem (np. Crick i Dodge, 1996; Graham i in., 1992; Wilkowski i Robinson; 2008). Jest to jedna z podstawowych ludzkich emocji (Berkowitz, 2012), która może być doświadczana jako stan przejściowy albo względnie stała cecha (Bąk, 2016; Spielberger, 2006). Na gniew rozumiany jako stan (*state anger*) składają się subiektywne odczucia oraz pobudzenie autonomicznego układu nerwowego, z kolei cecha gniewu (*trait anger*) to względnie trwała dyspozycja do reagowania złością w różnych sytuacjach (Spielberger i in., 1995). Wśród osób o wysokim poziomie cechy gniewu prawdopodobieństwo wystąpienia zachowań agresywnych zwiększa się, szczególnie w przypadku prowokacji (Bettencourt i in., 2006).

Ludzie różnią się pod względem wrażliwości na czynniki mogące wywołać zachowania agresywne (Lawrence, 2006). Osoby wrażliwe na prowokację mają skłonność do interpretowania sytuacji ambiwalentnych, trudnych do zinterpretowania jako wrogich (Green i Phillips, 2004). Powszechnie uważa się, że więźniowie charakteryzują się podwyższonym poziomem cechy gniewu i wrażliwości na prowokację, jednakże dotychczasowe wyniki badań dostarczyły niejednoznacznych dowodów w tym zakresie. Dlatego też w ramach badań własnych analizowałam między innymi to, czy

kobiety i mężczyźni, przebywający w polskich zakładach karnych, różnią się pod względem cechy gniewu i wrażliwości na prowokację **(artykuł 3).** Dotychczas nie zbadano także, czy wrażliwość na prowokację, związana z gniewem i wrogimi atrybucjami (Bondu i Richter, 2016), przekłada się na tendencje atrybucyjne w rozpoznawaniu emocji. Badania opisane w **artykule 4** uzupełniają tę lukę w literaturze.

4.3. Relacyjne korelaty wrogich atrybucji

Osobowość, którą można rozumieć jako zbiór struktur wiedzy, w tym schematów poznawczych (Allen i in., 2018), ma duże znaczenie dla sposobu, w jaki ludzie odbierają rzeczywistość i jak się zachowują w różnych sytuacjach społecznych. Jednak, jak podkreślają autorzy modelu GAM (Anderson i Bushman, 2002), w wyjaśnianiu przyczyn zachowań agresywnych ważną rolę odgrywają nie tylko zmienne osobowościowe, ale także sytuacyjne. Podobnie jest w przypadku odkrywania czynników związanych z wrogimi atrybucjami. Wykazano, że sposób interpretowania sytuacji społecznych zależy od kontekstu społeczno-relacyjnego (Zajenkowska i in. 2018; Zajenkowska i Rajchert, 2020). Ważne wydaje się więc uwzględnianie tego kontekstu w badaniach dotyczących związków wrogich atrybucji ze zmiennymi osobowymi, szczególnie z tymi, które wiążą się z problemami w relacjach z innymi ludźmi. Przykładem takiej cechy może być narcyzm wrażliwy.

Istnieją przynajmniej dwa rodzaje narcyzmu rozumianego jako cecha osobowości: narcyzm wrażliwy i wielkościowy (Dickinson i Pincus, 2003; Miller i Campbell, 2008; Wink, 1991). Obydwie formy wiążą się z poczuciem wyjątkowości, obydwie mogą także prowadzić do zachowań agresywnych. Jednakże ostatnie badania wykazały, że tylko narcyzm wrażliwy (pod którym kryje się m.in. niska samoocena i neurotyczność) jest związany z wrogim stylem atrybucyjnym (Hansen-Brown i Freis, 2021). Osoby o podwyższonym poziomie narcyzmu wrażliwego są skoncentrowane na

zapobieganiu potencjalnym niebezpieczeństwom, dlatego interpretując przyczyny zachowań innych ludzi, zakładają najgorsze i są gotowe do obrony (Freis, 2018; Hansen-Brown i Freis, 2021). Do tej pory nie zbadano jednak, na ile związek ten jest stabilny w zależności od sytuacji, dlatego też w **artykule 5** odpowiadam na to pytanie.

Uwzględnienie czynników relacyjnych jest ważne z jeszcze jednego powodu, mianowicie to właśnie relacje znacząco przyczyniają się do rozwoju stylu atrybucyjnego (Dodge, 2006). Sposób, w jaki kodujemy i interpretujemy docierające do nas informacje, zależy od wcześniej nabytych schematów (Crick i Dodge, 1994; Dodge, 2006). Według Kennetha A. Dodga (2006) tendencja do dokonywania wrogich atrybucji kształtuje się w dzieciństwie na podstawie nabytej wiedzy i doświadczenia. Jednakże kierunek tego „nabywania" nie jest oczywisty, bowiem reaktywne zachowania agresywne są naturalne i uniwersalne, pojawiają się w pierwszym roku życia człowieka w odpowiedzi np. na frustrację, a podstawowym zadaniem socjalizacji jest nauczenie się bycia nieagresywnym (Dodge, 2006). Podobnie jest w przypadku dokonywania wrogich atrybucji, początkowo dzieci dopasowują intencje do efektu działania. „Intencje muszą być zgodne z wynikiem" (Dodge, 2006; s. 793), a więc negatywne konsekwencje zachowania oznaczają negatywne intencje. Z czasem jednak większość dzieci zaczyna zauważać, że często ludzie działają „niespecjalnie" i – o ile nie pojawią się dowody na intencjonalne działanie aktora – dzieci przypisują łagodne atrybucje. Niektórzy jednak nie nabywają łagodnego wzorca atrybucyjnego. Przemoc, wrogość rodziców i rówieśników, niepowodzenia w ważnych doświadczeniach życiowych, wychowanie w kulturze honoru, pozabezpieczny styl przywiązania – wszystkie te czynniki mogą sprzyjać utrzymywaniu się wrogiej wizji świata (Dodge, 2006). Wydaje się, że zasadne jest dodanie do tej listy także cech osobowości rodziców, takich jak wspomniany już narcyzm wrażliwy, który może przełożyć się na sposób, w jaki rodzic odbiera rzeczywistość, i na jakość relacji z

dzieckiem (Rappoport, 2005). Między innymi w celu potwierdzenia tego przypuszczenia przeprowadziłam badania opisane w **artykule 6**.

W relacji kształtuje się także zdolność do mentalizowania, czyli rozumienia i adekwatnego interpretowania własnych i cudzych stanów psychicznych, takich jak: myśli, uczucia, potrzeby, pragnienia, przekonania i cele (Allen i in., 2008; Fonagy i in., 1991). Zdolność do mentalizowania jest częściowo wrodzona, a częściowo nabywana dzięki interakcji dziecka z obiektem przywiązania, którym najczęściej jest rodzic. Jeżeli opiekun adekwatnie odzwierciedla stany psychiczne niemowlęcia, jest uważny na jego pragnienia i potrzeby, dziecko uczy się regulować swoje emocje i prawidłowo odczytywać stany psychiczne (Bateman i Fonagy, 2008). Niektórym rodzicom jednak, na przykład tym z podwyższonym poziomem narcyzmu wrażliwego, może być trudno dostrzec emocje dziecka i właściwie je odzwierciedlić, co sprawia, że mały człowiek ma utrudnione warunki do rozwijania własnej zdolności do mentalizowania (Gergely, 2002). Deficyty te mogą w przyszłości przyczynić się m.in. do rozwoju tendencji do dokonywania wrogich atrybucji i agresji (Fonagy i Levinson, 2004; Zajenkowska i in., 2018), dlatego w **artykule 7** opisuję interwencję psychologiczną, której celem jest poprawienie zdolności do mentalizowania i redukcja wrogości w interpretowaniu rzeczywistości społecznej.

4.4. Wrogie atrybucje w różnych grupach

Początkowo badania dotyczące wrogich atrybucji prowadzono głównie wśród chłopców w młodym wieku, z czasem jednak zauważono, że konstrukt ten jest związany z agresją reaktywną także w przypadku osób dorosłych (Tuente, 2019). Wśród uczestników z grup nieklinicznych (dorosłych z populacji ogólnej, studentów) różnice indywidulane dotyczące stylu atrybucyjnego pojawiały się głównie w sytuacjach niejasnych, ambiwalentnych (Hutchings i in., 2010; Matthews i Norris, 2002). Im wyższy

(samoopisowy) poziom agresji, tym więcej wrogich atrybucji (Tuente, 2019). Mniej klarowne wyniki przedstawiono w pracach, w których autorzy porównywali poziom wrogich atrybucji wśród osób z populacji ogólnej i populacji klinicznej (pacjentów, osób osadzonych) – chociaż należy zauważyć, że tego typu badań wciąż nie ma wiele (Tuente, 2019). Generalnie uważa się, że osoby przebywające w więzieniach przejawiają wyższy poziom agresywności i chociaż także w tej grupie potwierdzono dodatnią korelację pomiędzy wrogimi atrybucjami a agresją (np. Edwards i Bond, 2012; Hornsveld i in., 2017), niełatwo znaleźć dowody naukowe, które wskazywałyby, że więźniowie są znacząco bardziej wrodzy niż osoby z populacji generalnej. Co więcej, dostępne wyniki badań potwierdzają, że osadzeni nie dokonują szczególnie dużej liczby wrogich atrybucji (Bowen i in., 2016; Zajenkowska i in., 2013, 2018; Zajenkowska, Rogoza i in., 2021).

Jeszcze więcej wątpliwości pojawia się, kiedy weźmiemy pod uwagę płeć. Niektóre prace kwestionują adekwatność używania modelu przetwarzania informacji społecznych w odniesieniu do dziewcząt (Crain i in., 2005). Nie ma wielu badań, w których testowano by różnice płciowe w zakresie wrogich atrybucji i agresji. Problem ten opisano w opracowaniach naukowych dotyczących zarówno osób dorosłych (Tuente, 2019), jak i dzieci (Martinelli i in., 2018). Najgorzej sytuacja wygląda w odniesieniu do osób osadzonych, gdzie właściwie cała dotychczasowa wiedza została oparta na badaniach przeprowadzonych wśród mężczyzn. Dlatego też w badaniach własnych (prowadzonych wraz z zespołem badawczym), jako jedni z nielicznych w Europie, postanowiliśmy przyjrzeć się dokonywaniu wrogich atrybucji przez kobiety skazane za przestępstwa z użyciem przemocy. Ponadto wykorzystaliśmy różnorodne metody pomiaru, aby zbadać zarówno sposób kodowania, jak i interpretacji niejednoznacznych wskazówek społecznych.

5. Główne wyniki badań własnych

Niniejsza rozprawa doktorska opiera na się na siedmiu artykułach opublikowanych w międzynarodowych (6 artykułów) i polskich (1 artykuł) czasopismach naukowych. Dodatkowo uwzględnia wyniki badań, które zaprezentowałam w trakcie międzynarodowej konferencji organizowanej przez The International Society for Research on Aggression (ISRA), która odbyła się w lipcu 2022 roku w Kanadzie[1]. Pracując wraz z zespołem badawczym, przeprowadziłam kilka badań, których wyniki przyczyniły się do pogłębienia wiedzy na temat czynników związanych z wrogimi atrybucjami. Badania te prowadziłam głównie w grupach osób mających problemy z reakcjami agresywnymi. Punktem wyjścia dla moich dociekań naukowych był model GAM i SIP. W związku z tym badania własne w dużej mierze skupiały się wokół czynników poznawczych i osobowościowych związanych z wrogimi atrybucjami.

5.1. Linia badań dotycząca czynników poznawczych

W pierwszych dwóch pracach (artykuły 1 i 2), będących efektem współpracy z zespołem badawczym, którego jestem członkinią (APSychoLAB), a także badaczami z zagranicy, przyglądaliśmy się sposobom kodowania niejednoznacznych sygnałów społecznych przez kobiety i mężczyzn skazanych za przestępstwa z użyciem przemocy w porównaniu z celowo dobranymi dorosłymi z populacji ogólnej. W dotychczasowych badaniach tego typu brali udział niemal wyłącznie sprawcy płci męskiej, wnioski zwykle generalizowano na wszystkie osoby osadzone, tymczasem – jak pokazują m.in. nasze wyniki badań – często była to błędna praktyka. Więźniowie i więźniarki różnią się pod wieloma względami, także w przypadku przetwarzania informacji społecznych, co wydaje się mieć znaczenie, na przykład przy tworzeniu programów resocjalizacyjnych.

[1] Otrzymałam stypendium ISRA2022 Student Travel Award.

W pracy *Attention orienting to the eyes in violent female and male offenders: An eye-tracking study* **(artykuł 1)** zbadaliśmy sposób, w jaki osoby skazane za przestępstwa z użyciem przemocy oraz osoby z populacji ogólnej patrzą na twarze wyrażające różne, niejednoznaczne emocje. Na uwagę zasługuje fakt, że jest to prawdopodobnie pierwsze tego typu badanie w Europie, uwzględniające sprawców obu płci i przeprowadzone na tak dużej próbie (N = 173). Artykuł ten powstał we współpracy z badaczami z Uniwersytetu w Tybindze, których prace mogłam obserwować w trakcie wizyty studyjnej, którą odbyłam w czerwcu 2022 roku. W omawianym badaniu wykorzystaliśmy technikę eyetrackingu oraz zadanie komputerowe składające się z serii fotografii przedstawiających zmorfowane twarze. Zdjęcia te powstały dzięki „wymieszaniu" trzech wyrazów emocjonalnych: gniewu, strachu i radości, w różnych proporcjach (np. 50% gniewu i 50% strachu). Okazało się, że zarówno osadzone kobiety, jak i mężczyźni zwracali mniejszą uwagę na oczy w początkowej fazie patrzenia na twarz. Dodatkowo analizy eksploracyjne ujawniły, że osoby, które wykazywały obniżoną uwagę na oczy, doświadczyły większych trudności w rozpoznawaniu emocji strachu.

W kolejnym artykule, pt. *Reduced attention toward faces, intentionality and blame ascription in violent offenders and community-based adults: Evidence from an eye-tracking study* **(artykuł 2)**, ponownie skupiliśmy się na wzorcach uwagowych osób skazanych za przestępstwa z użyciem przemocy (N = 63) oraz dobranej celowo grupy spośród populacji ogólnej (N = 63). Jednak tym razem postanowiliśmy sprawdzić po pierwsze, w jakim zakresie więźniowie w porównaniu z osobami niebędącymi więźniami zwracają uwagę na twarz aktorów przedstawionych w różnych sytuacjach, w których jedna jest ofiarą, a druga sprawcą określonego działania, i po drugie, który z przedstawionych bohaterów jest głównym przedmiotem zainteresowania. Dodatkowo w artykule tym zaproponowaliśmy nowy wskaźnik wrogich atrybucji – **izomorfizm**

intencjonalności i winy. Stanowi on odpowiedź na pytanie, które pojawiało się w naszym zespole od dawna. Wydaje się, że osoby skazane za przestępstwa z użyciem przemocy powinny charakteryzować się wyższym poziomem wrogich atrybucji niż osoby, które nigdy nie dokonały przestępstwa, tymczasem wcześniejsze badania systematycznie pokazywały odwrotną zależność – więźniowie mają tendencję do deklarowania niższej wrogości przy interpretowaniu niejednoznacznych sytuacji społecznych (Zajenkowska i in., 2013, 2018; Zajenkowska i in., 2021). W związku z czym postawiliśmy następującą hipotezę: kluczowy nie jest sam poziom przypisywanej intencjonalności i winy, ale raczej stopień ich zróżnicowania. Izomorfizm jest niczym innym jak zakresem nakładania się / korelacji pomiędzy dostrzeganą intencjonalnością zachowania i obwinianiem sprawcy za dane zachowanie. Przewidywaliśmy, że im wyższa korelacja pomiędzy tymi dwoma konstruktami (wyższy izomorfizm), tym bardziej automatyczna i nieelastyczna ocena sytuacji a tym samym wyższe prawdopodobieństwo wystąpienia gniewu oraz agresji. Wyniki tego badania potwierdziły nasze przypuszczenia – osoby osadzone charakteryzował wyższy izomorfizm intencjonalności i winy w porównaniu z osobami przebywającymi na wolności. Co więcej, okazało się, że osoby osadzone w mniejszym stopniu zwracają uwagę na twarze przedstawionych postaci. Jeżeli jednak już na nie patrzą, to podobnie jak osoby niebędące więźniami – skupiają się głównie na twarzach sprawców, a nie ofiar, co jest zgodne z ewolucyjnym wyjaśnieniem, mówiącym o tym, że bodźce zagrażające przyciągają uwagę.

Powyższe wyniki badań wzbogacają zbiór wiedzy na temat wzorców poznania związanych z wrogimi atrybucjami. Brak uważności na twarze wśród osób skłonnych do agresji może skutkować trudnościami w adekwatnym rozumieniu sytuacji społecznych, co prowadzi do aktywizacji wrogich schematów i ujawnienia się różnic indywidualnych.

5.2. Linia badań dotycząca czynników osobowościowych

Celem badania opisanego w artykule pt. *Sex Differences in Inmates: Anger, Sensitivity to Provocation and Family History of Imprisonment* **(artykuł 3)** było porównanie kobiet i mężczyzn skazanych za przestępstwa z użyciem przemocy (N = 123) oraz dorosłych przebywających na wolności (N = 118) w zakresie cechy gniewu i wrażliwości na prowokację, a także historii uwięzienia członków rodzin osób badanych. Wyniki wykazały, że kobiety (zarówno osadzone, jak i przebywające na wolności) deklarowały wyższą wrażliwość na prowokację niż mężczyźni. Nie zaobserwowano natomiast istotnych statystycznie różnic pomiędzy kobietami i mężczyznami pod względem cechy gniewu. Wyniki te w dużej mierze są zgodne z wcześniejszymi badaniami (np. Zajenkowska, 2014; Archer & Haigh, 1997) i mogą być interpretowane w kontekście ról społecznych – wydaje się, że kobiety są bardziej „uwrażliwione" na wszelkie oznaki gniewu i agresji, a jednocześnie w większym stopniu niż mężczyźni tłumią te emocje i reakcje.

Ponadto w naszym badaniu okazało się, że osoby z populacji ogólnej deklarowały wyższy poziom gniewu niż osoby osadzone. Ten zaskakujący wynik można interpretować co najmniej na dwa sposoby. Po pierwsze jest możliwe, że osoby przebywające w więzieniu doświadczyły desensytyzacji – przyzwyczaiły się do zdarzeń, które mogą wywoływać gniew. Po drugie, osoby osadzone mogą być szczególnie zmotywowane do tego, aby pokazywać się „w jak najlepszym świetle", w związku z czym przy prowadzeniu badań wśród więźniów ważne wydaje się uwzględnianie różnorodnych (nie tylko kwestionariuszowych) metod pomiaru.

W badaniu tym pytaliśmy także o krewnych, którzy zostali skazani na karę pozbawienia wolności. W grupie więźniów znacznie więcej osób miało krewnego, który także przebywał w zakładzie karnym, niż w grupie osób z populacji ogólnej. W

przypadku osadzonych kobiet krewnym tym najczęściej był ojciec lub matka, natomiast w przypadku mężczyzn – był to ojciec lub brat. Żaden z badanych mężczyzn odbywających karę pozbawienia wolności nie wskazał, że tego typu karę odbywała także jego matka. Uzyskane wyniki można odnieść do teorii społecznego uczenia się (Bandura, 1973), według której dzieci nabywają określone zachowania poprzez obserwację i naśladowanie zachowań rodziców, ale także w oparciu o wspomniany wcześniej model Dodga (2006), który odnosi się do rozwoju wrogich atrybucji.

Z kolei w pracy zatytułowanej *Gender differences in sensitivity to provocation and hostile attribution bias toward ambiguous facial cues in violent offenders and community-based adults* **(artykuł 4)** udzieliliśmy odpowiedzi na pytanie, czy wrażliwość na prowokację przewiduje tendencję do interpretowania niejednoznacznych ekspresji emocjonalnych twarzy jako rozgniewanych (wskaźnik wrogich atrybucji; Gehrer i in., 2019) i czy jest to zależne od płci osoby badanej i osoby przedstawionej na zdjęciu. W badaniu wzięło udział 105 więźniów i 167 osób, które nigdy nie zostały skazane. Ponownie wykorzystaliśmy zadanie składające się ze zmorfowanych twarzy, jednakże tym razem interesowała nas interpretacja emocji przedstawionych na zdjęciach. Osoby badane przyglądały się zmorfowanym twarzom przez krótki czas, a następnie odpowiadały na pytanie dotyczące emocji, jaką dostrzegały na obrazku. Zgodnie z hipotezą podatność na prowokację była związana z częstszym rozpoznawaniem gniewu na zmorfowanych męskich twarzach, jednak przede wszystkim w grupie mężczyzn, którzy nie byli skazani za przestępstwa agresywne. Natomiast osadzeni mężczyźni, charakteryzujący się wysokim poziomem wrażliwości na prowokację , widzieli więcej gniewu w twarzach kobiecych, ale mniej w twarzach męskich. Z kolei w przypadku osadzonych kobiet zaobserwowano następującą zależność – im wyższa wrażliwość na prowokację, tym niższa tendencja do interpretowania gniewu w kobiecych twarzach.

Badanie to pokazało, jak ważne jest uwzględnianie różnic międzypłciowych. Ta sama cecha może wiązać się z różnymi strategiami i sposobami interpretowania wskazówek dotyczących twarzy, a co za tym idzie, prowadzić do różnych zachowań w przypadku kobiet i mężczyzn. Ponadto, biorąc pod uwagę także wyniki opisane w artykule 3, można spekulować, że sama cecha gniewu czy podatności na prowokację nie różnicuje osób skłonnych do agresji od tych, które takiej skłonności nie przejawiają. Wyniki kolejnych badań sugerują, iż specyficzny kontekst społeczno-relacyjny, jak również cecha narcyzmu wrażliwego, wydają się mieć szczególne znaczenie w przypadku wrogich atrybucji.

5.3. Linia badań dotycząca czynników relacyjnych

W pracy pt. *Hostile attribution biases in vulnerable narcissists depends on the socio-relational context* **(artykuł 5)** zweryfikowałam związek narcyzmu wrażliwego z podwyższonym poziomem wrogich atrybucji w zależności od kontekstu społeczno-relacyjnego. W badaniu wzięły udział osoby z grupy klinicznej – pacjenci uczestniczący w psychoterapii grupowej (N = 74) oraz osoby z grupy nieklinicznej (N = 153) - ochotnicy, którzy zgłosili się do udziału w badaniu. W obydwu grupach badaliśmy zarówno poziom narcyzmu wrażliwego, jak i narcyzmu wielkościowego. Wszyscy uczestnicy wypełnili także kwestionariusz AIHQ (Ambiguous Intentions and Hostility Questionnaire; Combs i in., 2006). Jest to narzędzie służące do pomiaru wrogiego stylu interpersonalnego, składające się z opisu pięciu niejednoznacznych sytuacji, w każdej z nich występuje inna osoba: znajomy z pracy, ktoś ważny, ktoś obcy, znajomy i przyjaciel. Badanych poproszono, aby wyobrazili sobie, że to oni znaleźli się w danej sytuacji.

Dodatkowo w grupie nieklinicznej wzbogaciliśmy pomiar wrogich atrybucji o jeszcze jedno zadanie, które polegało na oglądaniu i interpretowaniu scen wizualnych wyświetlanych na monitorze komputera (Wilkowski i in., 2007; Zajenkowska i Rajchert, 2020). Wykorzystaliśmy trzy rodzaje obrazków: wrogie (wszystkie wskazówki umieszczone na ilustracji świadczyły o zamierzonym działaniu aktora np. wrogi wyraz twarzy, skierowanie ręki w stronę ofiary), ambiwalentne (niektóre elementy wskazywały na to, że działanie aktora było zamierzone, inne świadczyły o przypadkowości zdarzenia) oraz niewrogie (wszystkie wskazówki świadczyły o przypadkowym działaniu aktora). Osoby badane interpretowały oglądane obrazki z perspektywy obserwatora – osoby trzeciej, która nie bierze bezpośredniego udziału w zdarzeniu.

Wyniki naszego badania potwierdziły wcześniejsze doniesienia (Hansen-Brown i in., 2021) dotyczące pozytywnego związku wrogich atrybucji z narcyzmem wrażliwym (ale nie z narcyzmem wielkościowym). Zależność ta pojawiła się zarówno w grupie klinicznej, jak i nieklinicznej. Wykazaliśmy jednak, że prawidłowość ta zależy od kontekstu społeczno-relacyjnego, a także od poziomu ambiwalencji przedstawianych sytuacji. W obydwu grupach, w przypadku niejednoznacznych sytuacji społecznych, ocenianych z perspektywy własnej, w których brała udział osoba określona jako „przyjaciel", związek pomiędzy poziomem narcyzmu wrażliwego a wrogimi atrybucjami okazał się nieistotny. Podobnie w sytuacji z osobą określoną jako „autorytet", jednakże w tym przypadku brak związku odnotowano jedynie w grupie klinicznej. Co ciekawe, w zadaniu ze scenami wizualnymi pozytywna korelacja pomiędzy narcyzmem wrażliwym a wrogimi atrybucjami ujawniła się jedynie w sytuacjach zupełnie niewrogich, co może świadczyć o tym, że osoby o podwyższonym poziomie narcyzmu wrażliwego mają tendencję do dostrzegania wrogości tam, gdzie jej w rzeczywistości nie ma. Powyższy wynik, ale także doświadczenia z terapeutycznej pracy z dziećmi oraz

informacje uzyskane dzięki przeglądowi dostępnej literatury, sprawiły, że zaczęłam się zastanawiać nad znaczeniem podwyższonego poziomu narcyzmu wrażliwego w relacji rodzic – dziecko.

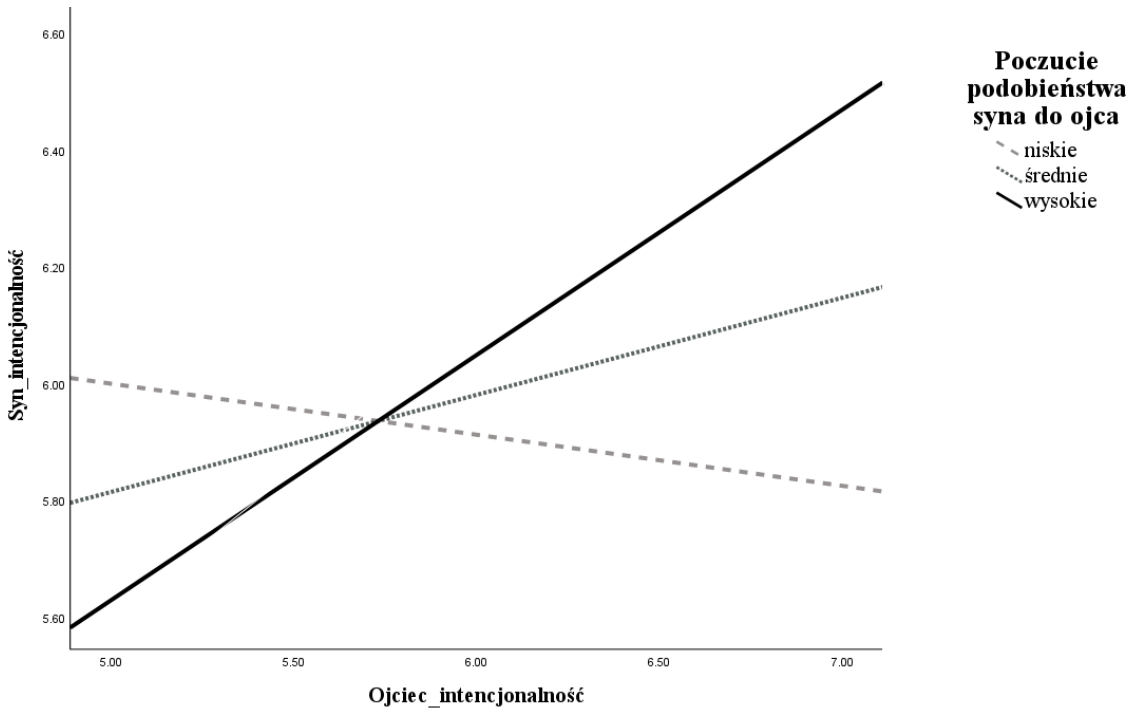
Wiedząc, że przekazywanie uwięzienia (w grupie osób skazanych za przestępstwa z użyciem przemocy) jest najbardziej widoczne wśród mężczyzn, których ojciec także przebywał w zakładzie karnym (np. Bodecka-Zych, 2021; Farrington, 2001), dostrzegając znaczenie narcyzmu wrażliwego we wrogim interpretowaniu rzeczywistości oraz znając teorię Dodga (2006), dotyczącą rozwoju wrogich atrybucji, postanowiłam zaprojektować badanie, które pozwoliłoby na udzielenie odpowiedzi na następujące pytania: czy poziom wrogich atrybucji ojca koreluje z poziomem wrogich atrybucji jego nastoletniego syna oraz czy znaczenie dla tej relacji ma poziom narcyzmu wrażliwego ojca. Wyniki zostały opisane w artykule pt. *Dad, are they laughing at me? Fathers' vulnerable narcissism and sons' hostile attribution* **(artykuł 6)**. Badanie to miało charakter diadyczny – objęło 77 par ojców i ich synów. Zastanawiając się nad metodologią, zauważyłam, że w większości wcześniejszych prac testujących zależność pomiędzy stylem artybucyjnym rodziców i dzieci, dorośli interpretowali zupełnie inne sytuacje niż ich synowie czy córki. Postanowiłam więc stworzyć autorskie narzędzie, oparte o wcześniejsze, zweryfikowane metody, które pozwoliłoby na uspójnienie pomiaru. W związku z czym procedura badawcza była następująca: podczas spotkania z nastolatkiem badacz czytał opisy trzech sytuacji, które potencjalnie mogłyby spotkać młodego człowieka. Dalej chłopcy interpretowali sytuacje, a na końcu proszono ich o opisanie dwóch wydarzeń ze swojego życia. Ojców natomiast proszono o wyobrażenie sobie, że sytuacje spotkały ich synów i byli zapoznawani ze wszystkimi pięcioma sytuacjami (tymi przytoczonymi przez badacza i tymi opisanymi przez ich syna). W przypadku każdego ojca sprawdzaliśmy poziom narcyzmu wrażliwego, natomiast

w przypadku syna – poziom agresywności (bycie zarówno ofiarą, jak i sprawcą). Wyniki analizowano oddzielnie dla każdego subfaktora wrogich atrybucji (intencjonalności, obwiniania i gniewnych uczuć). Okazało się, że nie ma bezpośredniego związku pomiędzy poziomem wrogich atrybucji ojców i synów. Jednakże poziom narcyzmu wrażliwego ojca koreluje pozytywnie z obwinianiem i uczuciami gniewu ujawnianymi przez syna. Dodatkowo narcyzm wrażliwy ojca przewiduje gniewne uczucia pojawiające się w odpowiedzi na działania innych osób, a te z kolei są związane z agresywnością syna (zarówno z byciem sprawcą, jak i ofiarą agresji). Wyniki tego badania sugerują, że nabywanie wrogich atrybucji może być związane z jakością relacji ojca z synem, dla której znaczenie może mieć poziom narcyzmu wrażliwego ojca.

Dodatkowo w projekcie dotyczącym przekazywania wrogich atrybucji z ojca na syna uwzględniłam dodatkową zmienną, jaką jest postrzegane podobieństwo osobowościowe syna do ojca (artykuł w przygotowaniu). Wcześniejsze badania sugerują, że poczucie podobieństwa do rodzica sprzyja identyfikacji z obiektem i wzmacnia proces modelowania (Bandura, 1969), co więcej, jeśli syn jest podobny do ojca, mężczyzna może być bardziej pewny swojego ojcostwa, co skutkuje większym zaangażowaniem w wychowanie dziecka (Gallup i in., 2016). W badaniu tym zapytałam synów, czy uważają, że ich zainteresowania, zachowania i postawy są podobne do zainteresowań, zachowań i postaw ich ojców. Następnie zarówno synowie, jak i ojcowie brali udział w zadaniu z wykorzystaniem scenek wizualnych (opis powyżej), tym razem jednak zadanie składało się wyłącznie z obrazków przedstawiających scenki ambiwalentne, a uczestników poproszono o wyobrażenie sobie, że to oni są jednym z przedstawionych aktorów (tym, któremu coś się dzieje). Okazało się, że postrzegane osobowościowe podobieństwo syna do ojca moderuje związek pomiędzy intencjonalnością przypisywaną przez ojców i przez synów (rys. 4.). Synowie, postrzegający siebie jako podobnych do

swojego taty pod względem osobowościowym, w zbliżony sposób przypisują intencjonalność w niejednoznacznych sytuacjach społecznych.

Wyniki dwóch wyżej opisanych badań zostały zaprezentowane na międzynarodowej konferencji organizowanej przez The International Society for Research on Aggression (ISRA), która odbyła się w lipcu 2022 roku w Kanadzie. Zostały zauważone i docenione m.in. przez Kennetha Dodga i Brada Bushmana – czołowych badaczy agresji i wrogich atrybucji, autorów modeli teoretycznych, na których opieram swoje badania.



Rysunek 4. Przypisywanie intencjonalności przez ojców i synów a poczucie podobieństwa osobowościowego syna do ojca, opracowanie własne.

5.4. Mentalizacja

Wielu badaczy zastanawia się, czy istnieje teoria psychologiczna, jakiś konstrukt, łączący te wszystkie elementy, które zaczynają się wyłaniać, kiedy przyglądamy się w sposób pogłębiony agresji i wrogim atrybucjom. Wydaje się, że odpowiedzą na to pytanie jest mentalizacja (Allen i in., 2008; Zajenkowska i in., 2021). Zaburzenia w zakresie mentalizowania mogą stanowić teoretyczne uzasadnienie dla większości wyników badań opisanych w niniejszej rozprawie. Przykładowo, więźniowie przejawiają szczególne deficyty w zakresie mentalizowania (Fonagy i Levinson, 2004). Można przypuszczać, że część z nich wychowywała się w rodzinach, które nie umożliwiły im nabycia tej kompetencji. Deficyty dotyczące dostrzegania i rozumienia stanów psychicznych mogą ujawniać się już na poziomie kodowania wskazówek społecznych. Jak pokazują nasze badania, osoby przebywające w więzieniu zwracają mniejszą uwagę na oczy i twarze – być może jest to mechanizm pozwalający na unikanie odczytywania stanów psychicznych innych ludzi. Dzieciom, które nie doświadczyły poczucia bezpieczeństwa w dzieciństwie łatwiej jest „wyłączyć" mentalizowanie, niż zauważyć, że dorosły czyni im krzywdę (Fonagy i Bateman, 2006). Co więcej, nieprawidłowa relacja z opiekunem i wyłączanie się struktur mózgowych odpowiedzialnych za mentalizowanie może powodować mniej elastyczne myślenie, które z kolei może być odpowiedzialne za wysoki izomorfizm intencjonalności i winy w grupie sprawców skazanych za przestępstwa z użyciem przemocy.

Wydaje się, że w pełni rozwinięta zdolność do mentalizowania może być „szczepionką" na wrogie atrybucje, dlatego w ostatnim artykule **(artykuł 7)** opisuję metodę, której celem jest redukcja tendencji do dokonywania wrogich atrybucji a tym samym chronienie przed impulsywnym, agresywnym działaniem. Mowa o treningu psychoedukacyjnym z elementami mentalizacji. Warsztat ten, oparty o skrypt stworzony

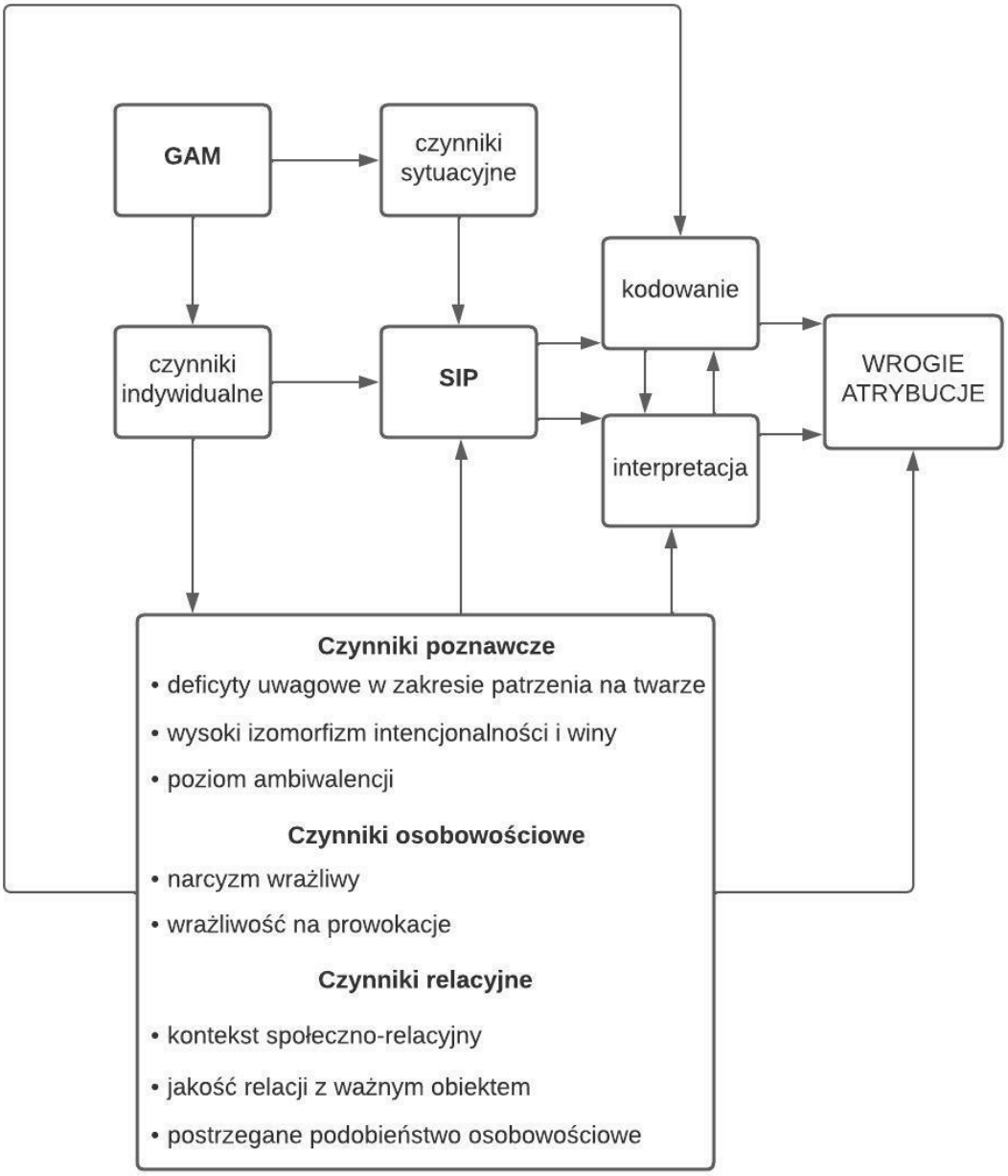
przez Batemana i Fonagy'ego (2013), poprawia zdolności mentalizacyjne – pomaga zrozumieć stany umysłowe własne i innych ludzi, a tym samym może pomóc korygować automatyczne, wrogie interpretacje (Zajenkowska i in., 2018). W artykule opublikowanym w polskim czasopiśmie „Psychologia Wychowawcza" zawracam uwagę na podstawy teoretyczne warsztatu, uzasadniam, dlaczego rozwijanie zdolności do mentalizowania jest kluczowe w przypadku osób osadzonych, opisuję także przebieg warsztatu z uwzględnieniem spostrzeżeń osób go prowadzących (liderów).

6. Konkluzje

Wnioski płynące z badań opisanych w tej pracy dostarczają nowych i istotnych informacji na temat wrogich atrybucji, stanowią także element polemiki naukowej, bowiem nie zawsze są zgodne z wynikami uzyskanymi wcześniej. Dzięki zastosowaniu zróżnicowanych metod i technik badawczych oraz uwzględnieniu odmiennych grup (np. kobiet skazanych za przestępstwa z użyciem przemocy, pacjentów biorących udział w terapii grupowej, par ojców i synów) ukazałam niuanse poznawczych, osobowościowych i relacyjnych korelatów wrogich atrybucji. Ponadto poszerzyłam wiedzę na temat wzorców kodowania i interpretowania sytuacji społecznych przez osoby doświadczające problemów z agresywnymi reakcjami. Zweryfikowałam też powszechnie przyjęte założenia dotyczące związków cech osobowości z tendencją do dokonywania wrogich atrybucji. Podkreśliłam również znaczenie kontekstu sytuacyjnego i rodzinnego dla działania i powstawania wrogiego stylu atrybucyjnego (rys. 5).

Oczywiście opisane tutaj badania mają swoje ograniczenia, głównym jest ich korelacyjny charakter. Na podstawie dostępnych opracowań teoretycznych możemy stawiać hipotezy o potencjalnych kierunkach związków, jednakże wciąż potrzebujemy kolejnych badań empirycznych. Dlatego w aktualnie realizowanym grancie Preludium zaplanowałam eksperymenty, które pozwolą między innymi na weryfikację znaczenia

postrzeganego podobieństwa osobowościowego syna do ojca oraz kontynuowanie badań dotyczących skuteczności treningu psychoedukacyjnego z elementami mentalizacji oraz rozpoczęcie badań dotyczących treningu wrażliwości na ekspresję emocjonalną. Trening ten działa na poziomie kodowania i polega na uczeniu badanych dostrzegania subtelnych zmian w ekspresjach emocjonalnych twarzy poprzez systematyczne wykonywanie zadań przy użyciu komputera. Oprócz tego realizuję badania jakościowe, które pozwolą na pogłębienie wiedzy na temat relacji synów z ojcami, którzy przebywają w zakładach karnych, a także uwzględniają perspektywę matek. Ostatecznym celem moich badań jest stworzenie kompleksowego modelu wyjaśniającego mechanizm nabywania skłonności do dokonywania wrogich atrybucji. Mam nadzieję, że moje badania znajdą także zastosowanie w praktyce – przyczynią się do tworzenia bardziej adekwatnych programów profilaktycznych i resocjalizacyjnych, których celem będzie redukcja wrogich atrybucji.



Rysunek 5. Czynniki poznawcze, osobowościowe i relacyjne związane z wrogimi atrybucjami w kontekście modelu GAM i SIP oraz zmienne zidentyfikowane w badaniach własnych, mające znaczenie dla powstawania i działania wrogich atrybucji, opracowanie własne.

Bibliografia

Allen, J. G., Fonagy, P., & Bateman, A. W. (2008). *Mentalizing in clinical practice*. American Psychiatric Pub.

Allen, J. J., & Anderson, C. A. (2017). Aggression and violence: Definitions and distinctions. *The Wiley handbook of violence and aggression*, 1-14.

Allen, J. J., Anderson, C. A., & Bushman, B. J. (2018). The general aggression model. *Current opinion in psychology*, *19*, 75-80.

Anderson, C. A., & Bushman, B. J. (2002). Human aggression. *Annual review of psychology*, *53*(1), 27-51.

Archer, J., & Haigh, A. (1997). Beliefs about aggression among male and female prisoners. *Aggressive Behavior, 23*(6), 405–415

Bąk, W. (2016). Pomiar stanu, cechy, ekspresji i kontroli złości. Polska adaptacja kwestionariusza Staxi-2. *Polskie Forum Psychologiczne*, 21, 93–122.

Bateman, A., & Fonagy, P. (2008). 8-year follow-up of patients treated for borderline personality disorder: mentalization-based treatment versus treatment as usual. *American Journal of Psychiatry*, *165*(5), 631-638.

Bateman, A., & Fonagy, P. (2013). Mentalization-based treatment. *Psychoanalytic inquiry*, *33*(6), 595-613.

Berkowitz, L. (2012). A different view of anger: The cognitive-neoassociation conception of the relation of anger to aggression. *Aggressive behavior*, *38*(4), 322-333.

Bettencourt, B., Talley, A., Benjamin, A. J., & Valentine, J. (2006). Personality and aggressive behavior under provoking and neutral conditions: a meta-analytic review. *Psychological bulletin*, 132(5), 751.

Blair, R. J. R. (2001). Neurocognitive models of aggression, the antisocial personality disorders, and psychopathy. Journal of Neurology, *Neurosurgery & Psychiatry*, 71(6), 727–731.

Bondü, R., & Richter, P. (2016). Interrelations of justice, rejection, provocation, and moral disgust sensitivity and their links with the hostile attribution bias, trait anger, and aggression. *Frontiers in Psychology*, 7, 795.

Bowen, K. N., Roberts, J. J., & Kocian, E. J. (2016). Decision making of inmates: Testing social information processing concepts using vignettes. *Applied Psychology in Criminal Justice*, *12*(1), 1-17.

Bushman, B. J., & Huesmann, L. R. (2010). Aggression. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of Social Psychology* (5th ed., Vol. 2, pp. 833–863). Hoboken, NJ: John Wiley & Sons.

Chapman, H., Gillespie, S. M., & Mitchell, I. J. (2018). Facial affect processing in incarcerated violent males: A systematic review. *Aggression and Violent Behavior, 38*, 123–138.

Combs, D. R., Penn, D. L., Wicher, M., & Waldheter, E. (2007). The Ambiguous Intentions Hostility Questionnaire (AIHQ): a new measure for evaluating hostile social-cognitive biases in paranoia. *Cognitive neuropsychiatry*, *12*(2), 128-143.

Crain, M. M., Finch, C. L., & Foster, S. L. (2005). The relevance of the social information processing model for understanding relational aggression in girls. *Merrill-Palmer Quarterly*, 51(2), 213–249.

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological bulletin*, *115*(1), 74.

Crick, N. R., & Dodge, K. A. (1996). Social information-processing mechanisms in reactive and proactive aggression. *Child development*, *67*(3), 993-1002.

De Castro, B. O., Veerman, J. W., Koops, W., Bosch, J. D., & Monshouwer, H. J. (2002). Hostile attribution of intent and aggressive behavior: A meta-analysis. *Child development*, *73*(3), 916-934.

DeWall, C. N., Anderson, C. A., & Bushman, B. J. (2012). Aggression. In H. Tennen, J. Suls, & I. B. Weiner (Eds.), *Handbook of psychology* (2nd ed., Vol. 5, pp. 449–466). Hoboken, NJ: John Wiley & Sons.

Dickinson, K. A., & Pincus, A. L. (2003). Interpersonal analysis of grandiose and vulnerable narcissism. Journal of Personality Disorders, 17(3), 188–207.

Dodge, K. A. (1980). Social cognition and children's aggressive behavior. *Child development*, *51*(1), 162-170.

Dodge, K. A. (2006). Translational science in action: Hostile attributional style and the development of aggressive behavior problems. *Development and psychopathology*, *18*(3), 791-814.

Dousari, A. S., Moghadam, M. T., & Satarzadeh, N. (2020). COVID-19 (Coronavirus disease 2019): a new coronavirus disease. *Infection and drug resistance*, 13, 2819.

Edwards, R., & Bond, A. J. (2012). Narcissism, self-concept clarity and aggressive cognitive bias amongst mentally disordered offenders. *Journal of Forensic Psychiatry & Psychology*, *23*(5-6), 620-634.

Epps, J., & Kendall, P. C. (1995). Hostile attributional bias in adults. *Cognitive Therapy and Research*, *19*(2), 159-178.

Farrington, D. P., Jolliffe, D., Loeber, R., Stouthamer-Loeber, M., & Kalb, L. M. (2001). The concentration of offenders in families, and family criminality in the prediction of boys' delinquency. *Journal of adolescence*, 24(5), 579-596.

Fonagy, P., & Bateman, A. (2006). Progress in the treatment of borderline personality disorder. *The British Journal of Psychiatry*, *188*(1), 1-3.

Fonagy, P., & Levinson, A. (2004). Offending and attachment: The relationship between interpersonal awareness and offending in a prison population with psychiatric disorder. *Canadian Journal of Psychoanalysis*, *12*(2), 225-251.

Fonagy, P., Steele, M., Steele, H., Moran, G. S., & Higgitt, A. C. (1991). The capacity for understanding mental states: The reflective self in parent and child and its significance for security of attachment. *Infant mental health journal*, *12*(3), 201-218.

Freis, S. D. (2018). The distinctiveness model of the narcissistic subtypes (DMNS): What binds and differentiates grandiose and vulnerable narcissism. In A. Brunnell, J. Foster, & T. Hermann (Eds.), *The handbook of trait narcissism: Key advances, research methods, and controversies* (pp. 37–46). Cham, Switzerland: Springer.

Gehrer, N. A., Duchowski, A. T., Jusyte, A., & Schönenberg, M. (2020). Eye contact during live social interaction in incarcerated psychopathic offenders. *Personality Disorders: Theory, Research, and Treatment*, 11(6), 431.

Gehrer, N. A., Sch¨onenberg, M., Duchowski, A. T., & Krejtz, K. (2018). *Implementing innovative gaze analytic methods in clinical psychology: A study on eye movements in antisocial violent offenders.* Paper Presented at the Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications.

Gergely, G., Fonagy, P., Jurist, E., & Target, M. (2002). Affect regulation, mentalization, and the development of the self. *International Journal of Psychoanalysis*, *77*, 217-234.

Graham, S., Hudley, C., & Williams, E. (1992). Attributional and emotional determinants of aggression among African-American and Latino young adolescents. *Developmental Psychology*, *28*(4), 731.

Green, M. J., & Phillips, M. L. (2004). Social threat perception and the evolution of paranoia. *Neuroscience & Biobehavioral Reviews*, *28*(3), 333-342.

Halligan, S. L., Cooper, P. J., Healy, S. J., & Murray, L. (2007). The attribution of hostile intent in mothers, fathers and their children. *Journal of abnormal child psychology*, *35*(4), 594-604.

Hansen-Brown, A. A., & Freis, S. D. (2021). Assuming the worst: Hostile attribution bias in vulnerable narcissists. *Self and Identity*, *20*(2), 152-164.

Hornsveld, R. H., Nijman, H. L., Hollin, C. R., & Kraaimaat, F. W. (2007). An adapted version of the Rosenzweig Picture-Frustration Study (PFS-AV) for the measurement of hostility in violent forensic psychiatric patients. *Criminal Behaviour and Mental Health*, *17*(1), 45-56.

Horsley, T. A., de Castro, B. O., & Van der Schoot, M. (2010). In the eye of the beholder: Eye-tracking assessment of social information processing in aggressive behavior. *Journal of abnormal child psychology*, *38*(5), 587-599.

Huesmann, L. R., & Taylor, L. D. (2006). The role of the mass media in violent behavior. In R. C. Brownson (Ed.), *Annual review of public health* (Vol. 26). Palo Alto, CA: Annual Reviews.

Hutchings, J. N., Gannon, T. A., & Gilchrist, E. (2010). A preliminary investigation of a new pictorial method of measuring aggression-supportive cognition among young aggressive males. *International journal of offender therapy and comparative criminology*, *54*(2), 236-249.

Lawrence, C. (2006). Measuring individual responses to aggression-triggering events: Development of the situational triggers of aggressive responses (STAR) scale. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, 32(3), 241-252.

Marsh, A. A., & Blair, R. J. R. (2008). Deficits in facial affect recognition among antisocial populations: A meta-analysis. *Neuroscience & Biobehavioral Reviews*, 32(3), 454–465.

Martinelli, A., Ackermann, K., Bernhard, A., Freitag, C. M., & Schwenck, C. (2018). Hostile attribution bias and aggression in children and adolescents: A systematic literature review on the influence of aggression subtype and gender. *Aggression and violent behavior*, *39*, 25-32.

Matthews, B. A., & Norris, F. H. (2002). When is believing "seeing"? Hostile attribution bias as a function of self-reported aggression 1. *Journal of Applied Social Psychology*, *32*(1), 1-31.

Miller, J. D., Hoffman, B. J., Gaughan, E. T., Gentile, B., Maples, J., & Keith Campbell, W. (2011). Grandiose and vulnerable narcissism: A nomological network analysis. *Journal of personality*, *79*(5), 1013-1042.

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.

Rappoport, A. (2005). Co-narcissism: How we accommodate to narcissistic parents. *The Therapist*, 1, 1–8.

Schönenberg, M., & Jusyte, A. (2014). Investigation of the hostile attribution bias toward ambiguous facial cues in antisocial violent offenders. *European archives of psychiatry and clinical neuroscience*, *264*(1), 61-69.

Slutkin, G., Ransford, C., & Decker, R. B. (2015). Cure violence: Treating violence as a contagious disease. In Maltz, M. D., & Rice, S. K. (Eds.), *Envisioning criminology* (pp. 43–56). Cham: Springer.

Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and decoding facial expressions. *Psychological Science*, 16(3), 184–189.

Spielberger, C. D. (2006). Cross-cultural assessment of emotional states and personality traits. *European Psychologist*, 11(4), 297.

Spielberger, C., & Vag~ P. (1995). Test anxiety: A transactional process model. In C. Spielberger & P. Vagg (Eds.), *Test anxiety: Theory, assessment, and treatment* (pp. 3-14). Washington, DC: Taylor & Francis

Tuente, S. K., Bogaerts, S., & Veling, W. (2019). Hostile attribution bias and aggression in adults-a systematic review. *Aggression and violent behavior*, *46*, 66-81.

Wilkowski, B. M., & Robinson, M. D. (2008). The cognitive basis of trait anger and reactive aggression: An integrative analysis. *Personality and social psychology review*, *12*(1), 3-21.

Wilkowski, B. M., Robinson, M. D., Gordon, R. D., & Troop-Gordon, W. (2007). Tracking the evil eye: Trait anger and selective attention within ambiguously hostile scenes. *Journal of Research in Personality*, *41*(3), 650-666.

Wink, P. (1991). Two faces of narcissism. *Journal of personality and social psychology*, *61*(4), 590.

Young, J. E., Klosko, J. S., & Weishaar, M. E. (2003). *Schema therapy*. New York: Guilford.

Zajenkowska, A., & Rajchert, J. (2020). How sensitivity to provocation shapes encoding and interpretation of ambivalent scenes in an eye tracking study. *Journal of Cognitive Psychology*, *32*(2), 180-198.

Zajenkowska, A., Bower Russa, M., Rogoza, R., Park, J., Jasielska, D., & Skrzypek, M. (2021). Cultural influences on social information processing: Hostile attributions in the United States, Poland, and Japan. *Journal of Personality Assessment*, *103*(4), 489-497.

Zajenkowska, A., Jankowski, K. S., Lawrence, C., & Zajenkowski, M. (2013). Personality and individual differences in responses to aggression triggering events among prisoners and non-prisoners. *Personality and Individual Differences*, *55*(8), 947-951.

Zajenkowska, A., Mylonas, K., Lawrence, C., Konopka, K., & Rajchert, J. (2014). Cross cultural sex differences in situational triggers of aggressive responses. *International Journal of Psychology*, *49*(5), 355–363.

Zajenkowska, A., Prusik, M., & Szulawski, M. (2018). What does the Ambiguous Intentions Hostility Questionnaire really measure? The importance of context in evaluating hostility bias. *Journal of Personality Assessment, 102*(2), 205–213.

Zajenkowska, A., Prusik, M., Jasielska, D., & Szulawski, M. (2021). Hostile attribution bias among offenders and non-offenders: Making social information processing more adequate. *Journal of Community & Applied Social Psychology*, *31*(2), 241-256.

Zajenkowska, A., Rogoza, R., Sasson, N. J., Harvey, P. D., Penn, D. L., & Pinkham, A. E. (2021). Situational context influences the degree of hostile attributions made by individuals with schizophrenia or autism spectrum disorder. *British Journal of Clinical Psychology*, *60*(2), 160-176.

Artykuły składające się na cykl pracy doktorskiej oraz oświadczenia współautorów

Biological Psychology 163 (2021) 108136

Contents lists available at ScienceDirect

Biological Psychology

journal homepage: www.elsevier.com/locate/biopsycho

Attention orienting to the eyes in violent female and male offenders: An eye-tracking study

Nina A. Gehrer [a,*], Anna Zajenkowska [b], Marta Bodecka [b], Michael Schönenberg [a,c]

[a] University of Tübingen, Department of Clinical Psychology and Psychotherapy, Tübingen, Germany
[b] Maria Grzegorzewska University, Department of Psychology, Warsaw, Poland
[c] University Hospital Tübingen, Department of Psychiatry and Psychotherapy, Tübingen, Germany

ARTICLE INFO

Keywords:
Female offenders
Violent offenders
Eye gaze
Attention to the eyes
Eye tracking

ABSTRACT

Attention to the eyes and eye contact form an important basis for the development of empathy and social competences including prosocial behavior. Thus, impairments in attention to the eyes of an interaction partner might play a role in the etiology of antisocial behavior and violence. For the first time, the present study extends investigations of eye gaze to a large sample (N = 173) including not only male but also female violent offenders and a control group. We assessed viewing patterns during the categorization of emotional faces via eye tracking. Our results indicate a reduced frequency of initial attention shifts to the eyes in female and male offenders compared to controls, while there were no general group differences in overall attention to the eye region (i.e., relative dwell time). Thus, we conclude that violent offenders might be able to compensate for deficits in spontaneous attention orienting during later stages of information processing.

1. Introduction

A sensitivity to socially salient cues such as the face and the eyes of other individuals typically starts to develop during infancy (Farroni, Csibra, Simion, & Johnson, 2002; Haith, Bergman, & Moore, 1977). Already at the age of 3–6 months, infants respond with increased attention, smiling and facial mimicry to eye contact initiated by their parents (de Klerk, Hamilton, & Southgate, 2018; Hains & Muir, 1996). Eye contact is the basis for the development of 'intersubjectivity', which fosters adequate social interactions (Tronick, 1989). It is beneficial to attend to the eyes of interaction partners, since information from the eye region is crucial to draw conclusions regarding the mental or emotional state of a person (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001). Further, someone's gaze direction facilitates the understanding of their perceptions and intentions and, thus, helps to interpret their behavior (Brooks & Meltzoff, 2014; Emery, 2000; Klein, Shepherd, & Platt, 2009). Eye contact is considered an *ostensive cue*, which indicates that someone has the intent to communicate with us (Csibra & Gergely, 2006; Sperber & Wilson, 1986). Therefore, attention to the eyes of other individuals and maintaining eye contact are assumed to be essential factors for social development and particularly for the development of empathy (Dadds et al., 2012; Emery, 2000; Itier & Batty, 2009).

An empathic response includes the understanding of the other's mental and emotional state (cognitive empathy) as well as sharing the feeling (affective empathy; Eisenberg & Eggum, 2009; Hoffman, 2008). Empathy is a strong motive for prosocial behavior, since helping or comforting others can reduce the shared negative feelings (Hoffman, 2008). Further, prosocial behavior depends on the cognitive ability for explicit and reflective understanding that actions are driven by underlying mental states (i.e., beliefs, knowledge, percepts, intentions, etc.), which is also conceptualized as *theory of mind* or broader as *mentalizing* and represents mainly the cognitive component of empathy (Fonagy, Gergely, & Target, 2007; Frith, 2001; Leslie, 1987). Violence is related to a momentary inhibition of the capacity for mentalizing, i.e., for interpreting actions as being based on intentional mental states (Bateman & Fonagy, 2008). Therefore, antisocial and violent behavior seems to be linked to impairments in mentalizing and is often associated with callousness, a lack of empathy, and other affective and interpersonal impairments (Neumann & Hare, 2008). Based on these assumptions, early impairments in attention orienting to the eyes and deficient eye contact might compromise the development of cognitive as well as affective aspects of empathy and thereby contribute to the development of affective-interpersonal abnormalities and antisocial, deviant behavior (Dadds et al., 2012; Gehrer, Duchowski, Jusyte, & Schönenberg, 2020;

* Corresponding author at: University of Tübingen, Department of Clinical Psychology and Psychotherapy, Schleichstr. 4, D-72076, Tübingen, Germany.
E-mail address: nina.gehrer@uni-tuebingen.de (N.A. Gehrer).

https://doi.org/10.1016/j.biopsycho.2021.108136
Received 30 December 2020; Received in revised form 10 June 2021; Accepted 10 June 2021
Available online 12 June 2021

Waller & Hyde, 2018).

Furthermore, previous studies have indicated a recognition deficit for emotional expressions in violent and antisocial male offenders (Chapman, Gillespie, & Mitchell, 2018; Marsh & Blair, 2008). This recognition deficit – particularly regarding distress cues such as fearful faces – might also contribute to the development and maintenance of aggressive behavior and psychopathic traits (Blair, 1995, 2001). Reduced attention to the eyes in violent offenders could partly explain the impairments in emotion recognition, since the eyes have been argued to be a central diagnostic feature for emotional expressions (e.g., wide open fearful eyes; Smith, Cottrell, Gosselin, & Schyns, 2005; Whalen et al., 2004). However, findings regarding a relationship of viewing patterns and facial affect recognition are equivocal. A couple of studies in healthy samples report a connection between recognition accuracy of facial emotions and attention to diagnostic regions such as the eyes (Green & Guo, 2018; Pollux, Hall, & Guo, 2014; Schurgin et al., 2014; Vaidya, Jin, & Fellows, 2014), whereas the evidence in offender samples, to date, is limited (Dargis, Wolf, & Koenigs, 2018; Gehrer, Schönenberg, Duchowski, & Krejtz, 2018; Gehrer, Scheeff, Jusyte, & Schönenberg, 2019; Gillespie, Rotshtein, Beech, & Mitchell, 2017). Nevertheless, aforementioned results strengthen the assumed importance of possible impairments in attention to the eyes in offenders.

A growing body of research has investigated the association of reduced attention to the eyes (eye gaze) while viewing faces with patterns of antisocial and delinquent behavior and with affective and interpersonal impairments such as a lack of empathy, callousness or manipulative behavior (Billeci et al., 2019; Boll & Gamer, 2016; Bours et al., 2018; Dadds, El Masry, Wimalaweera, & Guastella, 2008; Dargis et al., 2018; Gehrer et al., 2018, 2019; Gillespie, Rotshtein, Wells, Beech, & Mitchell, 2015, 2017; Martin-Key, Graf, Adams, & Fairchild, 2018; Mowle, Edens, Ruchensky, & Penson, 2019; van Zonneveld, Platje, de Sonneville, van Goozen, & Swaab, 2017). Some studies have extended these investigations to eye contact during live social interaction (Dadds, Jambrak, Pasalich, Hawes, & Brennan, 2011, 2012; Dadds et al., 2014; Gehrer et al., 2020). Up to now, the investigations include children and adolescents with and without conduct problems (Billeci et al., 2019; Bours et al., 2018; Dadds et al., 2008; Martin-Key et al., 2018; van Zonneveld et al., 2017), adult community samples (Boll & Gamer, 2016; Gillespie et al., 2015; Mowle et al., 2019) and male offender groups (Dargis et al., 2018; Gehrer et al., 2018, 2019; Gehrer et al., 2020; Gillespie et al., 2017). Overall, there is evidence for deficient attention to the eyes associated with higher psychopathic traits (e.g., Dadds et al., 2008; Dargis et al., 2018; Gehrer et al., 2019; Gillespie et al., 2017, 2015; but see, e.g., Martin-Key et al., 2018; Mowle et al., 2019) and some of the studies indicate a link of the impairments particularly with affective psychopathic traits and its precursors in childhood and adolescence (i.e., a lack of empathy, callousness, a lack of remorse and guilt; Dadds et al., 2014, Dadds et al., 2008; Dadds et al., 2011; Gehrer et al., 2020). In contrast, the link between impaired eye gaze and behavioral aspects such as conduct problems during childhood and adolescence, criminal activities and antisocial behavior in general has received clearly less support (Billeci et al., 2019; Gehrer et al., 2018; Gillespie et al., 2017; Martin-Key et al., 2018; van Zonneveld et al., 2017).

To date, two studies have compared attention to the eyes in male violent offenders and non-offending controls (Gehrer et al., 2018; Gillespie et al., 2017). Both studies assessed eye movements during categorization of faces with different emotional expressions. The results indicated no group differences, neither in gaze metrics for early attention orienting nor in overall attention to the eye region. This may suggest that even if higher psychopathic traits are linked to reduced eye gaze within offender groups (Dargis et al., 2018; Gehrer et al., 2019, 2020; Gillespie et al., 2017), impairments in attention to the eyes are not associated with antisocial and criminal behavior in general. However, all these studies in offender samples share one fundamental limitation: they only included male participants. This might be based on lower numbers of women who are convicted of criminal offenses compared to men and therefore more limited access to study them (Suter, Byrne, Byrne, Howells, & Day, 2002). Yet, previous studies have reported gender differences in eye gaze (e.g., Hall, Hutton, & Morgan, 2010; Sullivan, Campbell, Hutton, & Ruffman, 2017) and in social attention in general that are already evident in infants (Christov-Moore et al., 2014; Rennels & Cummings, 2013). Typically, females exhibit higher levels of attention to the eyes compared to males which occurs across both genders of the displayed faces and across different emotional expressions (e.g., Hall et al., 2010; Martin-Key et al., 2018; Sullivan et al., 2017 but see Wells, Gillespie, & Rotshtein, 2016). Furthermore, several studies have reported a female advantage in emotion recognition and higher levels of empathy in females compared to males (Christov-Moore et al., 2014; Thompson & Voyer, 2014). It is likely that gender differences in viewing patterns are linked to the reported gender differences in emotion recognition and empathy, although findings are equivocal and the exact relationship still needs to be clarified (Christov-Moore et al., 2014; Cowan, Vanman, & Nielsen, 2014; Hall et al., 2010). Thus, we cannot simply assume that previous findings on eye gaze in male violent offenders also apply to female offenders. Therefore, it is crucial to extend previous research to female offenders to develop a full understanding of antisocial behavior and possible associations with attention to the eyes.

For the first time, the current study assessed attention orienting to the eyes in female as well as male violent offenders and provides a comparison to an age- and gender-matched control group. Further, a larger total sample size of $N = 173$ participants increases the probability to detect group differences with smaller effect size. Attention to the eyes was assessed during a categorization task of facial expressions. Since this study was part of a larger project with the aim to investigate the interpretation of ambiguous social stimuli in violent offenders, the emotional expressions varied in ambiguity. The current work mainly focuses on the least ambiguous conditions to enable an investigation of recognition accuracy of emotional faces and its link to attention to the eyes in this sample. The analysis included gaze metrics for early and general attention orienting to the eye region, i.e., the frequency of the initial fixation after stimulus onset and the relative dwell time during the stimulus presentation. With regard to general sex differences, we hypothesized that females compared to males would show higher levels of attention to the eyes as indicated by previous studies reporting sex differences in viewing patterns towards faces (Hall et al., 2010; Sullivan et al., 2017). Furthermore, aforementioned limited studies among male offenders lead to the hypothesis that violent offenders do not differ from non-offending controls. We wanted to test this prediction among male and female offenders. For this analysis, we aimed to consider possible group differences depending on the emotional expression (and the gender) of the displayed faces (please see supplements for an analysis of eye-tracking data considering the gender of the faces). Finally, we report the recognition accuracy and an explorative analysis of its relationship to attention to the eyes.

2. Materials and methods

2.1. Participants

We recruited four groups of participants, i.e., female and male incarcerated violent offenders and female and male healthy control participants (N = 227). After exclusion of datasets that yielded insufficient data quality (mean proportion of data samples per trial <75 %) and three participants that did not look at the images in more than 10 % of the trials, we included 82 incarcerated offenders (female: $N = 36$; male: $N = 46$) in our analysis and 91 healthy controls (female: $N = 36$; male: $N = 55$) of comparable age. The offender groups were recruited from two cooperating Polish correctional facilities in Siedlce and in Łódź. All of them had been convicted of violent crimes including murder, attempted murder, physical assault, etc. The participants of the community sample were recruited via social media or directly approached by the

investigators. All participants were aged between 22 and 61 years and native speakers of Polish. The study was approved by the university's ethics committee and conducted in accordance with the Declaration of Helsinki. All participants gave their informed consent prior to participation in the study.

2.2. Measures

We assessed the fluid intelligence of all participants using the 11 Matrix Reasoning items from the International Cognitive Ability Resource (ICAR; The International Cognitive Ability Resource Team, 2014). It is a public-domain tool for assessing cognitive abilities in psychology and social sciences. The Matrix Reasoning items contain stimuli that are comparable to those used in Raven's Progressive Matrices (Raven, Raven, & Court, 1998). Eleven 3 × 3 arrays of geometric shapes with one of the nine shapes missing are presented to the participant. Additionally, six geometric shapes are displayed as response alternatives and participants are asked to choose which one best completes the arrays. This instrument has been successfully used in previous research (Condon & Revelle, 2014; Dworak, Revelle, Doebler, & Condon, 2020) also in Poland (e.g., Karwowski & Brzeski, 2017).

2.3. Ambivalence task

The experimental paradigm was an extended version of the ambivalence task that has been used in previous offender studies (Schönenberg & Jusyte, 2014) but also in studies of patients with attention-deficit/hyperactivity disorder and social anxiety (Jusyte & Schönenberg, 2014; Schneidt, Jusyte, & Schönenberg, 2019). In this task, participants were asked to interpret ambiguous emotional expressions of facial images. The stimuli were created using pictures from the Radboud Faces Database (Langner et al., 2010). Pictures of six models (three females and three males; code 5, 7, 8, 27, 30, 57) with three different expressions (i.e., angry, fearful, and happy) were selected based on the recognition accuracy in the validation study (Langner et al., 2010). All 18 pictures were cropped and for each model the different emotional expressions were morphed into each other using FantaMorph software (Abrosoft, Beijing, China). This created three continuous affective dimensions for each model, i.e., angry-fearful, angry-happy, and fearful-happy. Finally, five ratios of the blended emotional expressions were exported from each continuum (i.e., 10:90 %, 30:70 %, 50:50 %, 70:30 %, and 90:10%) leading to a total of 90 stimuli (see Fig. 1 for the 15 stimuli of one of the female models). During the task all stimuli were presented twice (i.e., 180 trials). The female and male stimuli were divided in two blocks, since female stimuli have been added as extension compared to the previously designed task (Schönenberg & Jusyte, 2014). Within the two blocks, stimuli were presented in random order. The participants were asked to decide for each stimulus which emotion was displayed, i.e., angry, fearful, or happy.

2.4. Procedure and apparatus

Participants were tested in designated rooms of the correctional facilities or at the university. As this experiment was part of a greater project, participants were asked to do the ambivalence task and another computer task and to fill out paper questionnaires including the ICAR and some questions to assess demographics. First, they were asked to complete the ambivalence task while their eye movements were recorded by a Tobii Pro X3-120 remote eye tracker with a sampling rate of 120 Hz. A 9-point calibration was performed and repeated until a sufficient level of measurement accuracy was achieved (i.e., an average error of max. 0.68° visual angle). The data collection was monitored by an experimenter. Stimulus presentation and data recording was controlled by a Dell laptop via the iMotions Attention Tool software (version 7.2; iMotions A/S, Copenhagen, Denmark). The stimuli were presented on a 17'' computer screen (1920 × 1080 pixels resolution; stimulus size: ca. 8 × 10° visual angle), while the participants were seated at a viewing distance of approximately 60−70 cm. Each trial started with a presentation of a fixation cross in the upper left or bottom right corner of the screen (1000−2000 ms). A fixation on the cross location ensured an unbiased interpretation of the first fixation within the stimulus (Gehrer et al., 2019), since the start position at stimulus onset has been shown to influence following gaze patterns (Arizpe, Kravitz, Yovel, & Baker, 2012). Subsequently, the stimulus was presented between 500 and 1500 ms. The varying presentation times reduce the top-down driven influences on the gaze patterns due to decreased predictability of stimulus onset and offset. At the end of each trial, the response display showed the three possible responses (angry,

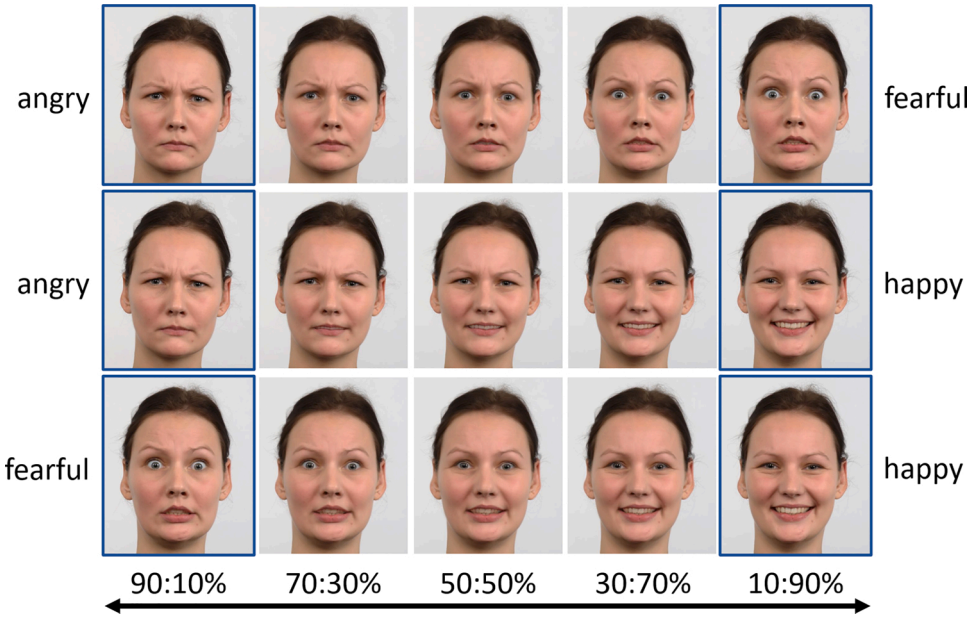


Fig. 1. Example of the stimuli for one female model. The face shows emotional expressions of different ambiguity on the three continua angry-fearful, angry-happy, and fearful-happy. The ratio of blended emotions is indicated in the bottom line. A blue frame highlights the conditions with the lowest ambiguity on which our main analysis was focused (see section on data analysis for more details).

fearful, and happy) that were presented until the participant selected one of the emotion categories via mouse click.

2.5. Data analysis

Since this article aims to investigate possible gender differences in early attention orienting towards the eyes of face stimuli in violent offenders, the analysis mainly focuses on the eye tracking data. Additionally, we display the results for the emotion recognition performance and investigate their relationship with attention to the eyes. The data analysis was conducted using the iMotions Attention Tool software (version 7.2; iMotions A/S, Copenhagen, Denmark) and R (version 3.5.2; R Core Team, 2018, Vienna, Austria). The eye region was defined as area of interest (AOI; a rectangle including both eyes, see Fig. 2) for each of the models and we computed the frequency of the initial fixation after stimulus onset being on this AOI as a measure of early attention orienting. Further, we calculated the dwell time on the eye region relative to the absolute dwell time on the stimulus (relative dwell time; RDT) as a measure of general attention. Before aggregating the data, we excluded 3.48 % of the trials due to technical problems such as poor data quality (i.e., missing data for >50 % of the trial). For the main analysis, we focused on trials with the lowest ambiguity, i.e., 90:10 % (and 10:90 %) ratio of blended emotions, to enable the investigation of possible effects of emotion type on attention to the eyes. Finally, we computed the means of the frequency of the initial fixation on the eye region and the RDT on the eye region for all emotional expressions and for each participant. Analyses of variance (ANOVAs) were used to test for main effects of the factors group (offenders vs. controls) and sex (female vs. male), emotional expression (angry, fearful, happy), and their two- and three-way interactions. Furthermore, we analyzed group and sex differences while considering the gender of the displayed faces across all emotional conditions and all levels of ambiguity in a supplemental analysis including not only the 90:10 % ratio of blended emotions but all conditions (see Supplements). As measures of effect size, we report partial eta squared (η_p^2) and Cohen's *d*.

For the analysis of emotion recognition accuracy, three participants were excluded that did not respond in more than 10 % of the trials (one female offender and two male offenders). As in the analysis of eye-tracking data, we focused on trials with the lowest ambiguity, i.e., 90:10 % ratio of blended emotions, and considered the 90 % emotion as correct response. The proportion of correct responses was analyzed implementing an ANOVA with the within-subjects factor emotion category (fearful, angry, happy) and with group and sex as between-subjects factors. Furthermore, associations between emotion recognition accuracy and attention to the eye region were analyzed by means of Pearson's product moment correlations separately for all emotion categories based on the particular importance of the eye region for fearful faces (Smith et al., 2005; Whalen et al., 2004).

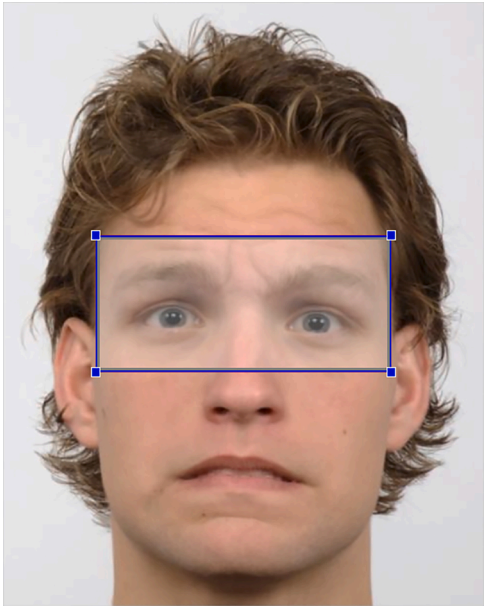


Fig. 2. Illustration of the definition of the AOI for a male face. This face represents a 50:50 % mixture of the emotional expressions angry and fearful.

3. Results

3.1. Participant characteristics

Demographic and diagnostic data of all four groups (offenders and controls; female and male participants) are displayed in Table 1. Group differences in age and intelligence were investigated by 2 (group) x 2 (sex) ANOVAs. The analysis revealed no significant age differences between the groups, all *F*s < 1.64, all *p*s > .202. However, intelligence differed significantly between groups, which was indicated by a significant main effect of group, $F(1, 169) = 32.64$, $p < .001$, $\eta_p^2 = .16$ and a significant interaction of group and sex, $F(1, 169) = 8.04$, $p = .005$, $\eta_p^2 = .05$. The main effect of sex remained non-significant, $F(1, 169) = 0.01$, $p = .907$, $\eta_p^2 = .00$. Numerically, male controls showed the highest intelligence scores, followed by female controls, female offenders, and finally male offenders. Post-hoc *t* tests revealed significant differences between most of the groups, all *t*s > 2.08, all *p*s < .041. Only the differences between female controls and female offenders or male controls were only marginally significant, all *t*s >1.80, all *p*s < .076. Furthermore, the education level was generally higher in control participants compared to offenders (see Table 1).

3.2. Eye-tracking data

Fig. 3 displays the frequency of the initial fixation on the eyes for both groups (offenders vs. controls) and both sexes (female vs. male participants) and for all emotion categories (angry, fearful, happy). The analysis for the frequency of the initial fixation on the eyes yielded a significant main effect of group (offenders vs. controls), $F(1, 169) = 4.68$, $p = .032$, $\eta_p^2 = .03$, indicating a reduced frequency of initial fixations on the eye region for the offenders ($M = 56.89$ %, $SD = 18.69$ %) compared to controls ($M = 62.48$ %, $SD = 18.45$ %). Further, the main effect of sex was marginally significant, $F(1, 169) = 2.86$, $p = .093$, $\eta_p^2 = .02$, indicating a tendency for a higher frequency of initial fixations on the eyes for male compared to female participants. Finally, the main

Table 1
Demographic and diagnostic sample characteristics.

	Female controls (*n* = 36)	Male controls (*n* = 55)	Female offenders (*n* = 36)	Male offenders (*n* = 46)
Age (years)	36.42 (11.05)	33.62 (10.04)	36.67 (9.73)	37.39 (9.91)
Intelligence (ICAR)*	4.89 (2.74)	6.00 (2.59)	3.81 (2.34)	2.78 (2.03)
Education level				
Primary	2.8 %	7.3 %	38.9 %	36.9 %
Secondary	27.8 %	30.9 %	30.6 %	47.8 %
High school	33.3 %	34.5 %	19.4 %	13.0 %
Higher	36.1 %	27.3 %	5.6 %	0 %
Other/not reported	0 %	0 %	5.6 %	0 %

Note. The data represented in the table refers to means and standard deviations (in parentheses) for age and intelligence and to relative frequencies for education level. ICAR = International Cognitive Ability Resource.

* For intelligence, post-hoc *t* tests indicate significant differences between all groups except for female controls who only differ from male offenders but not from male controls and female offenders.

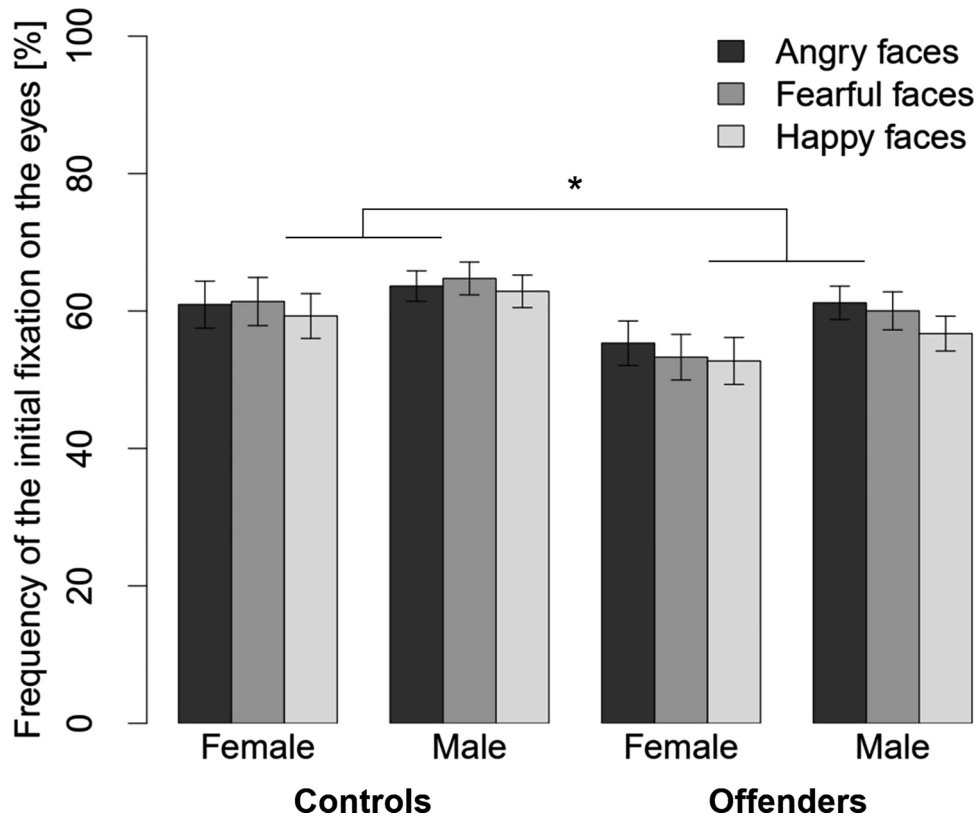


Fig. 3. Relative frequency of the initial fixation on the eyes for both groups (offenders and control participants) and both sexes (female and male) and for all emotion categories. Error bars represent standard error of the mean. The asterisk indicates the significant group difference.

effect of emotion category was marginally significant, $F(1.97, 333.25) = 2.76$, $p = .065$, $\eta_p^2 = .02$[1], whereas all interactions remained non-significant, all $Fs < 0.79$, all $ps > .451$. Additionally, we conducted an analysis of covariance (ANCOVA) including intelligence (ICAR) as covariate to control for possible effects of higher intelligence in offenders compared to controls. The influence of intelligence remained non-significant, $F(1, 168) = 0.49$, $p = .484$, $\eta_p^2 = .00$.

Fig. 4 displays the RDT on the eyes for both groups (offenders vs. controls) and both sexes (female vs. male participants) and for all emotion categories (angry, fearful, happy). The analysis for RDT on the eyes indicated a significant main effect of emotion category, $F(1.95, 329.10) = 31.71$, $p < .001$, $\eta_p^2 = .16$, while all other main effects and all interactions remained non-significant, all $Fs < 2.11$, all $ps > .148$. Pairwise post-hoc t tests comparing the RDT on the eyes for the different emotion categories indicated that RDT on the eyes was significantly reduced for happy faces compared to angry faces, $t(172) = 6.19$, $p < .001$, $d = 0.18$, and to fearful faces, $t(172) = 7.46$, $p < .001$, $d = 0.20$, while RDT on the eyes was similar for angry and fearful faces, $t(172) = 1.10$, $p = .274$, $d = 0.03$. Additionally, we conducted an ANCOVA including intelligence (ICAR) as covariate. The influence of intelligence was marginally significant, $F(1, 168) = 3.33$, $p = .070$, $\eta_p^2 = .02$. The test statistics of all other effects did not change considerably.

3.3. Behavioral data and associations

The emotion recognition accuracy overall was high ($M = 0.91$, $SD = 0.08$; for group- and emotion-specific means and standard deviations, see Table S1 in supplements) and the analysis yielded a significant main effect of emotion category, $F(1.89, 313.95) = 55.58$, $p < .001$, $\eta_p^2 = .25$. All other main effects and all interactions remained non-significant, all $Fs < 2.64$, all $ps > .106$. Post-hoc t tests comparing the proportion of correct responses for the three emotion categories revealed that all categories differed significantly, all $ts > 2.98$, all $ps < .003$. The recognition accuracy was the highest for happy faces ($M = 0.97$, $SD = 0.06$) followed by angry faces ($M = 0.89$, $SD = 0.14$) and fearful faces ($M = 0.86$, $SD = 0.14$). The analysis of associations between the accuracy and attention to the eyes revealed significant positive correlations between both measures of attention to the eyes and recognition accuracy for fearful faces, i.e., frequency of the initial fixation on the eyes: $r(168) = .21$, $p = .007$ and RDT on the eyes: $r(168) = .24$, $p = .002$. All correlations for happy and angry faces remained non-significant, all $|rs| < .03$, all $ps > .667$.

4. Discussion

In the current study, we investigated attention to the eyes for the first time not only in male but also in female violent offenders. Eye movements were recorded while viewing and categorizing emotional expressions of face stimuli. The results indicate generally reduced early attention shifts to the eyes in female as well as male offenders compared to female and male control participants of comparable age (i.e., lower frequency of the initial fixation on the eyes after stimulus onset). General attention to the eye region did not differ between offenders and controls. Furthermore, we did not replicate previously reported sex differences in attention to the eyes. The recognition accuracy was generally high and there were no group or sex differences. However, our

[1] Exploratory post-hoc t tests indicated that the marginally significant main effect of emotion category was driven by a reduced frequency of initial fixations on the eyes for happy compared to angry faces (significant) and fearful faces (marginally significant). Please note that this tendency is comparable to the results for RDT on the eyes. However, this must be interpreted with caution since the corresponding main effect of emotion was non-significant.

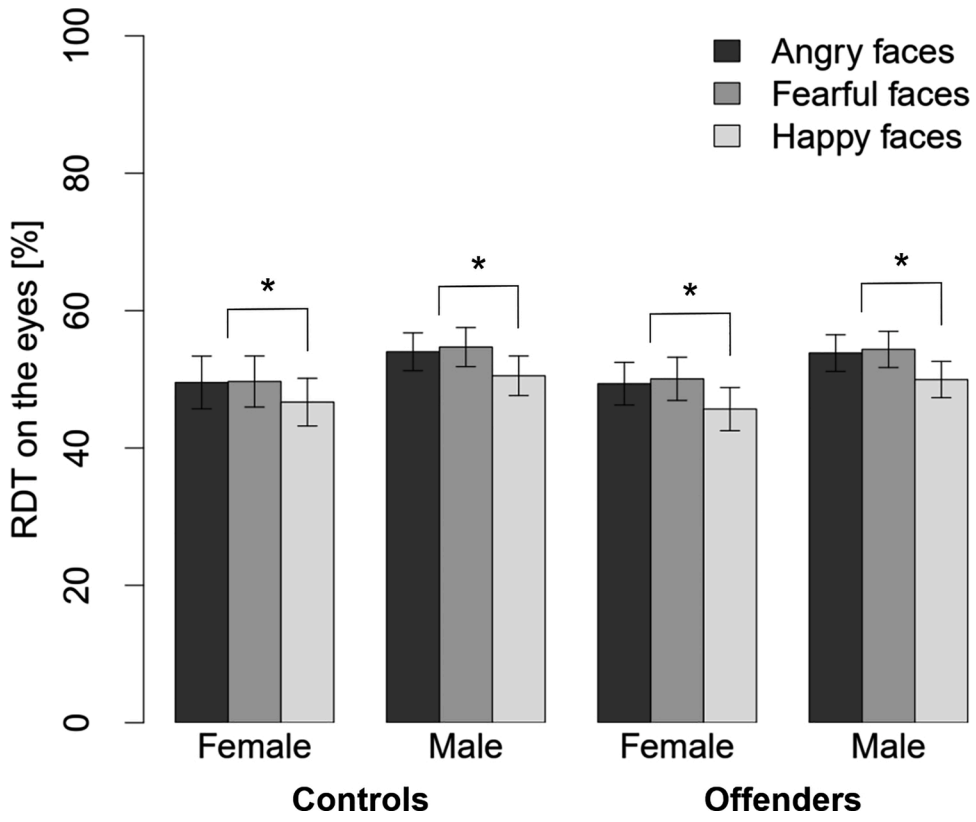


Fig. 4. RDT on the eyes for both groups (offenders and control participants) and both sexes (female and male) and for all emotion categories. Error bars represent standard error of the mean. The asterisk indicates the significant difference between happy compared to fearful and angry faces.

exploratory analysis of the relationship between attention to the eyes and recognition accuracy revealed positive associations of both measures of attention to the eyes and the proportion of correct responses specific for fearful faces. Therefore, individuals who exhibited reduced attention to the eyes were less accurate in the identification of fearful faces.

The reduced frequency of initial attention orienting to the eyes in female and male violent offenders occurred contrary to our hypothesis. Only two previous studies investigated attention to the eyes in male violent offenders compared to community samples (Gehrer et al., 2018; Gillespie et al., 2017). They studied comparable measures of overall and early attention orienting to the eyes (Gehrer et al., 2018: RDT and frequency of the initial fixation; Gillespie et al., 2017: dwell time, fixation count, and first fixation time) and both studies yielded no group differences. However, the conclusions based on these studies were limited due to relatively small sample sizes and the focus on male participants. Therefore, we aimed to re-investigate possible impairments in attention to the eyes in violent offenders in a larger sample while also including female participants. Overall, all groups (female and male offenders and control participants) exhibited a tendency to focus on the eye region: on average, over 50 % of initial fixations were directed to the eyes and the eye region was fixated for over 50 % of the stimulus presentation time. Further, our large total sample size (N = 173) allowed to detect a group difference in the frequency of initial attention shifts to the eyes between offenders and controls despite the small effect size. On average, the control group initially directed attention to the eye region in 62.48 % of the trials, while the offenders showed that only in 56.89 % of the trials. Since initial saccades have been demonstrated to be less influenced by goal-directed (top-down) processes (Nummenmaa, Hyönä, & Calvo, 2006), reduced spontaneous attention orienting to the eye region in offenders probably indicates differences in stimulus-driven (bottom-up) processes of attention orienting. In contrast, general attention orienting during the full stimulus presentation did not reflect the reduced early attention to the eye region exhibited by offenders. Therefore, we could speculate that violent offenders are able to compensate for small deficits in spontaneous attention orienting to the eyes by top-down processes in later information processing stages. Since the group difference is rather small this does not necessarily imply a role of deficient attention to socially important cues such as the eyes in the development and maintenance of violent and antisocial behavior. However, future studies should explore the importance of reduced spontaneous eye gaze in the etiology of violence further because as a potential risk factor for the development of violent behavior it might be a target for early intervention and prevention strategies (Dadds, English, Wimalaweera, Schollar-Root, & Hawes, 2019).

While, to date, the evidence for a relationship between impaired attention to the eyes and antisocial behavior in adults, adolescents and children has been weak (Billeci et al., 2019; Gehrer et al., 2018; Gillespie et al., 2017; Martin-Key et al., 2018; van Zonneveld et al., 2017), psychopathy and its precursors have been linked to reduced eye gaze by several studies (Dadds et al., 2008, 2011; Dadds et al., 2014; Dargis et al., 2018; Gehrer et al., 2019; Gillespie et al., 2015, 2017). Psychopathic traits are assumed to represent a dimensional construct (Edens, Marcus, Lilienfeld, & Poythress, 2006; Guay, Ruscio, Knight, & Hare, 2007), even if others suggest a conceptualization as taxon (psychopath vs. non-psychopath; DeLisi, 2016; Wright, 2009). Overall, previous research has indicated a link of impaired attention to the eyes particularly with affective psychopathic traits and its precursors in childhood and adolescence (i.e., a lack of empathy, callousness, a lack of remorse and guilt; Dadds et al., 2014, Dadds et al., 2008, 2011; Gehrer et al., 2020). In contrast to the rather small group differences in early attention to the eyes in the present study, a previous study has yielded clearly reduced eye gaze in offenders with high psychopathic traits compared to offenders with low psychopathic traits (Gehrer et al., 2019). These impairments associated with psychopathy affected early as well as overall attention to the eyes. Thus, reductions in eye gaze were more general

and more than six times larger than in the current study. Since, we did not assess psychopathic traits in the current study, we cannot exclude that reduced early attention shifts to the eyes in the offender sample compared to the control group might be explained by higher affective or interpersonal psychopathic traits instead of antisocial behavior. Furthermore, future studies could benefit from an additional assessment of empathy to extend the understanding of the connection between attention to the eyes, empathy, and antisocial and violent behavior.

Besides the missing assessment of psychopathic traits, our study has further limitations. First, intelligence was higher in the control group compared to offenders and therefore, should be considered when interpreting group differences. Even though previous studies (e.g., Gehrer et al., 2019) have not reported associations between intelligence and attention to the eyes, we investigated possible effects of intelligence in additional ANCOVAs. While the covariate intelligence was not significant for the frequency of initial fixations on the eye region, it reached marginal significance for RDT on the eyes. Thus, we assume that the reported group differences in early attention orienting to the eyes are not based on group differences in intelligence. However, we cannot rule out that intelligence might play a role in eye gaze, particularly with regard to processes after the initial orientation that are stronger influenced by top-down processes (Nummenmaa et al., 2006). Thus, future studies should investigate attention to the eyes in samples that are matched in intelligence and education.

Second, we could not replicate sex differences in eye gaze: While previous findings showed reduced initial and general attention to the eyes in male compared to female individuals (Hall et al., 2010; Martin-Key et al., 2018; Sullivan et al., 2017), male participants exhibited numerically higher levels of early attention orienting to the eyes in our study, although this effect did not reach significance. We conclude that gender differences in attention to the eyes might vary according to external factors such as stimulus, task, and setting as well as the relationship between eye gaze and empathy (compare Cowan et al., 2014; and Hall et al., 2010). Therefore, future research in violent female and male offenders needs to extend the investigation of attention to the eyes to other tasks and other settings (e.g., social interaction).

Another limitation of the current study is that we assessed eye movements during the categorization of emotional faces of varying ambiguity because this was part of a larger project investigating the interpretation of ambivalent social stimuli in violent offenders. To investigate emotion-specific effects on eye gaze and to calculate emotion recognition accuracy, our main analysis focused only on the trials with the lowest ambiguity. However, our supplementary analysis included all conditions and additionally considered the effects of the gender of the displayed faces. This analysis revealed the same main results with regard to group and sex of the participants with one exception (see supplements): Male controls tend to look longer at the eye region of female compared to male faces, whereas male offenders and females do not differentiate. Further, there was an overall tendency for more early attention shifts to the eyes of male compared to female faces. However, since we only presented images of three female and three male models, we cannot exclude that other gender-independent differences between the stimuli lead to this effect. Since effects of the gender of the displayed face and their interactions with emotional expression and intensity have been shown to be rather complex (see Wells et al., 2016 for a detailed discussion), future studies should include images of more female and male models to increase representativeness.

Finally, the presentation time of the stimuli was relatively short (500–1500 ms) compared to previous eye-tracking studies in offenders (i.e., 2000–2500 ms; Gehrer et al., 2018; Gillespie et al., 2017), which does not allow the investigation of dwell time over a longer period and thus may limit the comparability across studies. However, the shorter presentation time might have allowed the detection of the fear-specific positive associations of attention to the eyes and recognition accuracy, since early attention orienting is more important for the categorization if there is less time to view all parts of the face. This association strengthens the assumed importance of the eye region for the identification of fearful expressions (wide open fearful eyes; e.g., Smith et al., 2005; Whalen et al., 2004). Nevertheless, the reported associations were small ($r = .21–.24$) and group differences in attention to the eyes were not reflected in fear-specific or general emotion recognition impairments. However, the lack of an emotion recognition deficit in violent offenders is consistent with previous studies showing that this deficit can only be captured by more difficult tasks implementing not only full-blown but also more subtle emotional expressions (Gillespie et al., 2017; Schönenberg et al., 2014).

Besides the reported limitations, our study also presents several strengths. The current study is the first to extend the investigation of eye gaze to female violent offenders, which is crucial to develop a complete understanding of the characteristics of violent and antisocial offenders in general. Our findings show that female and male violent offenders do not substantially differ in attention orienting towards faces. Further, the large sample size allowed the detection of small group differences and the analysis of measures of early as well as general attention orienting allowed us to disentangle impairments in specific mechanisms of attention. Finally, our findings for general effects of emotion type on recognition accuracy and attention to the eyes are consistent with earlier findings. Accordingly, happy faces were across all groups identified with the highest accuracy and were associated with reduced overall attention to the eyes compared to fearful and angry faces (e.g., Gehrer et al., 2019; Schurgin et al., 2014).

5. Conclusion

Impairments in attention to the eyes and eye contact might facilitate the development of antisocial and delinquent behavior by compromising the development of empathy and social cognition, which form the basis for prosocial behavior. To date, the support for a possible association between reduced attention to the eyes and antisocial behavior has been weak and studies in violent offenders have been limited to male participants. The present study investigates attention to the eyes for the first time not only in male but also in female violent offenders. Overall, our findings indicate a small impairment in spontaneous attention shifts to the eye region in female and male offenders compared to the control group. Future studies need to clarify if these differences can be explained by higher psychopathic traits in the offender samples and if they are relevant for the etiology of violent behavior and therefore a reasonable target of intervention or prevention strategies.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgements

We would like to thank the staff of the correctional facilities in Siedlce and in Łódź for their support in recruitment of participants and organization of data collection. This study was funded by a grant to AZ from the National Science Centre (UMO-2017/26/D/HS6/00258).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.biopsycho.2021.108136.

References

Arizpe, J., Kravitz, D. J., Yovel, G., & Baker, C. I. (2012). Start position strongly influences fixation patterns during face processing: Difficulties with eye movements as a measure of information use. *PloS One, 7*(2), Article e31106. https://doi.org/10.1371/journal.pone.0031106

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "reading the mind in the eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines, 42*(2), 241–251. https://doi.org/10.1017/S0021963001006643

Bateman, A., & Fonagy, P. (2008). Comorbid antisocial and borderline personality disorders: Mentalization-based treatment. *Journal of Clinical Psychology, 64*(2), 181–194. https://doi.org/10.1002/jclp.20451

Billeci, L., Muratori, P., Calderoni, S., Chericoni, N., Levantini, V., Milone, A., & Dadds, M. (2019). Emotional processing deficits in Italian children with Disruptive Behavior Disorder: The role of callous unemotional traits. *Behaviour Research and Therapy, 113*, 32–38. https://doi.org/10.1016/j.brat.2018.12.011

Blair, R. J. R. (1995). A cognitive developmental approach to morality: Investigating the psychopath. *Cognition, 57*(1), 1–29. https://doi.org/10.1016/0010-0277(95)00676-P

Blair, R. J. R. (2001). Neurocognitive models of aggression, the antisocial personality disorders, and psychopathy. *Journal of Neurology, Neurosurgery & Psychiatry, 71*(6), 727–731. https://doi.org/10.1136/jnnp.71.6.727

Boll, S., & Gamer, M. (2016). Psychopathic traits affect the visual exploration of facial expressions. *Biological Psychology, 117*, 194–201. https://doi.org/10.1016/j.biopsycho.2016.03.010

Bours, C. C. A. H., Bakker-Huvenaars, M. J., Tramper, J., Bielczyk, N., Scheepers, F., Nijhof, K. S., … Buitelaar, J. K. (2018). Emotional face recognition in male adolescents with autism spectrum disorder or disruptive behavior disorder: An eye-tracking study. *European Child & Adolescent Psychiatry, 27*(9), 1143–1157. https://doi.org/10.1007/s00787-018-1174-4

Brooks, R., & Meltzoff, A. N. (2014). Gaze following: A mechanism for building social connections between infants and adults. *Mechanisms of Social Connection: From Brain to Group* (pp. 167–183).

Chapman, H., Gillespie, S. M., & Mitchell, I. J. (2018). Facial affect processing in incarcerated violent males: A systematic review. *Aggression and Violent Behavior, 38*, 123–138. https://doi.org/10.1016/j.avb.2017.10.006

Christov-Moore, L., Simpson, E. A., Coudé, G., Grigaityte, K., Iacoboni, M., & Ferrari, P. F. (2014). Empathy: Gender effects in brain and behavior. *Neuroscience & Biobehavioral Reviews, 46*, 604–627. https://doi.org/10.1016/j.neubiorev.2014.09.001

Condon, D. M., & Revelle, W. (2014). The international cognitive ability resource: Development and initial validation of a public-domain measure. *Intelligence, 43*, 52–64. https://doi.org/10.1016/j.intell.2014.01.004

Cowan, D. G., Vanman, E. J., & Nielsen, M. (2014). Motivated empathy: The mechanics of the empathic gaze. *Cognition & Emotion, 28*(8), 1522–1530. https://doi.org/10.1080/02699931.2014.890563

Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In *Processes of Change in Brain and Cognitive Development. Attention and Performance XXI* (Vol. 21, pp. 249–274).

Dadds, M. R., El Masry, Y., Wimalaweera, S., & Guastella, A. J. (2008). Reduced eye gaze explains "fear blindness" in childhood psychopathic traits. *Journal of the American Academy of Child & Adolescent Psychiatry, 47*(4), 455–463. https://doi.org/10.1097/CHI.0b013e31816407f1

Dadds, M. R., Allen, J. L., Oliver, B. R., Faulkner, N., Legge, K., Moul, C., & Scott, S. (2012). Love, eye contact and the developmental origins of empathy v. psychopathy. *British Journal of Psychiatry, 200*(3), 191–196. https://doi.org/10.1192/bjp.bp.110.085720

Dadds, M. R., Allen, J. L., McGregor, K., Woolgar, M., Viding, E., & Scott, S. (2014). Callous-unemotional traits in children and mechanisms of impaired eye contact during expressions of love: A treatment target? *Journal of Child Psychology and Psychiatry, 55*(7), 771–780. https://doi.org/10.1111/jcpp.12155

Dadds, M. R., English, T., Wimalaweera, S., Schollar-Root, O., & Hawes, D. J. (2019). Can reciprocated parent–child eye gaze and emotional engagement enhance treatment for children with conduct problems and callous-unemotional traits: a proof-of-concept trial. *Journal of Child Psychology and Psychiatry, 60*(6). https://doi.org/10.1111/jcpp.13023

Dadds, M. R., Jambrak, J., Pasalich, D., Hawes, D. J., & Brennan, J. (2011). Impaired attention to the eyes of attachment figures and the developmental origins of psychopathy. *Journal of Child Psychology and Psychiatry, 52*(3), 238–245. https://doi.org/10.1111/j.1469-7610.2010.02323.x

Dargis, M., Wolf, R. C., & Koenigs, M. (2018). Psychopathic traits are associated with reduced fixations to the eye region of fearful faces. *Journal of Abnormal Psychology, 127*(1), 43–50. https://doi.org/10.1037/abn0000322

de Klerk, C. C. J. M., Hamilton, A. F.d. C., & Southgate, V. (2018). Eye contact modulates facial mimicry in 4-month-old infants: An EMG and fNIRS study. *Cortex, 106*, 93–103. https://doi.org/10.1016/j.cortex.2018.05.002

DeLisi, M. (2016). Psychopathy, assessment, and the structure of crime. In M. DeLisi (Ed.), *Psychopathy as unified theory of crime* (pp. 61–95). New York: Palgrave Macmillan US.

Dworak, E. M., Revelle, W., Doebler, P., & Condon, D. M. (2020). Using the International Cognitive Ability Resource as an open source tool to explore individual differences in cognitive ability. *Personality and Individual Differences.* , Article 109906. https://doi.org/10.1016/j.paid.2020.109906

Edens, J. F., Marcus, D. K., Lilienfeld, S. O., & Poythress, N. G., Jr. (2006). Psychopathic, not psychopath: Taxometric evidence for the dimensional structure of psychopathy. *Journal of Abnormal Psychology, 115*(1), 131–144. https://doi.org/10.1037/0021-843X.115.1.131

Eisenberg, N., & Eggum, N. D. (2009). Empathic responding: Sympathy and personal distress. In *The social neuroscience of empathy* (Vol. 6, pp. 71–83).

Emery, N. J. (2000). The eyes have it: The neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews, 24*(6), 581–604. https://doi.org/10.1016/S0149-7634(00)00025-7

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences, 99*(14), 9602–9605. https://doi.org/10.1073/pnas.152159999

Fonagy, P., Gergely, G., & Target, M. (2007). The parent–infant dyad and the construction of the subjective self. *Journal of Child Psychology and Psychiatry, 48*(3–4), 288–328. https://doi.org/10.1111/j.1469-7610.2007.01727.x

Frith, U. (2001). Mind blindness and the brain in autism. *Neuron, 32*(6), 969–979. https://doi.org/10.1016/S0896-6273(01)00552-9

Gehrer, N. A., Duchowski, A. T., Jusyte, A., & Schönenberg, M. (2020). Eye contact during live social interaction in incarcerated psychopathic offenders. *Personality Disorders: Theory, Research, and Treatment*. https://doi.org/10.1037/per0000400

Gehrer, N. A., Scheeff, J., Jusyte, A., & Schönenberg, M. (2019). Impaired attention toward the eyes in psychopathic offenders: Evidence from an eye tracking study. *Behaviour Research and Therapy, 118*, 121–129. https://doi.org/10.1016/j.brat.2019.04.009

Gehrer, N. A., Schönenberg, M., Duchowski, A. T., & Krejtz, K. (2018). Implementing innovative gaze analytic methods in clinical psychology: A study on eye movements in antisocial violent offenders. *Paper Presented at the Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*.

Gillespie, S. M., Rotshtein, P., Beech, A. R., & Mitchell, I. J. (2017). Boldness psychopathic traits predict reduced gaze toward fearful eyes in men with a history of violence. *Biological Psychology, 128*, 29–38. https://doi.org/10.1016/j.biopsycho.2017.07.003

Gillespie, S. M., Rotshtein, P., Wells, L. J., Beech, A. R., & Mitchell, I. J. (2015). Psychopathic traits are associated with reduced attention to the eyes of emotional faces among adult male non-offenders. *Frontiers in Human Neuroscience, 9*(552), 1–11. https://doi.org/10.3389/fnhum.2015.00552

Green, C., & Guo, K. (2018). Factors contributing to individual differences in facial expression categorisation. *Cognition & Emotion, 32*(1), 37–48. https://doi.org/10.1080/02699931.2016.1273200

Guay, J.-P., Ruscio, J., Knight, R. A., & Hare, R. D. (2007). A taxometric analysis of the latent structure of psychopathy: Evidence for dimensionality. *Journal of Abnormal Psychology, 116*(4), 701–716. https://doi.org/10.1037/0021-843X.116.4.701

Hains, S. M. J., & Muir, D. W. (1996). Infant sensitivity to adult eye direction. *Child Development, 67*(5), 1940–1951. https://doi.org/10.1111/j.1467-8624.1996.tb01836.x

Haith, M. M., Bergman, T., & Moore, M. J. (1977). Eye contact and face scanning in early infancy. *Science, 198*(4319), 853–855. https://doi.org/10.1126/science.918670

Hall, J. K., Hutton, S. B., & Morgan, M. J. (2010). Sex differences in scanning faces: Does attention to the eyes explain female superiority in facial expression recognition? *Cognition & Emotion, 24*(4), 629–637. https://doi.org/10.1080/02699930902906882

Hoffman, M. L. (2008). Empathy and prosocial behavior. *Handbook of Emotions, 3*, 440–455.

Itier, R. J., & Batty, M. (2009). Neural bases of eye and gaze processing: The core of social cognition. *Neuroscience & Biobehavioral Reviews, 33*(6), 843–863. https://doi.org/10.1016/j.neubiorev.2009.02.004

Jusyte, A., & Schönenberg, M. (2014). Threat processing in generalized social phobia: An investigation of interpretation biases in ambiguous facial affect. *Psychiatry Research, 217*(1), 100–106. https://doi.org/10.1016/j.psychres.2013.12.031

Karwowski, M., & Brzeski, A. (2017). Selfies and the (creative) self: A diary study. *Frontiers in Psychology, 8*(172), 1–10. https://doi.org/10.3389/fpsyg.2017.00172

Klein, J. T., Shepherd, S. V., & Platt, M. L. (2009). Social attention and the brain. *Current Biology, 19*(20), R958–R962. https://doi.org/10.1016/j.cub.2009.08.010

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 & Emotion, 24*(8), 1377–1388. https://doi.org/10.1080/02699930903485076

Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind". *Psychological Review, 94*(4), 412–426. https://doi.org/10.1037/0033-295X.94.4.412

Marsh, A. A., & Blair, R. J. R. (2008). Deficits in facial affect recognition among antisocial populations: A meta-analysis. *Neuroscience & Biobehavioral Reviews, 32*(3), 454–465. https://doi.org/10.1016/j.neubiorev.2007.08.003

Martin-Key, N. A., Graf, E. W., Adams, W. J., & Fairchild, G. (2018). Facial emotion recognition and eye movement behaviour in conduct disorder. *Journal of Child Psychology and Psychiatry, 59*(3), 247–257. https://doi.org/10.1111/jcpp.12795

Mowle, E. N., Edens, J. F., Ruchensky, J. R., & Penson, B. N. (2019). Triarchic psychopathy and deficits in facial affect recognition. *Journal of Personality, 87*(2), 240–251. https://doi.org/10.1111/jopy.12386

Neumann, C. S., & Hare, R. D. (2008). Psychopathic traits in a large community sample: Links to violence, alcohol use, and intelligence. *Journal of Consulting and Clinical Psychology, 76*(5), 893–899. https://doi.org/10.1037/0022-006X.76.5.893

Nummenmaa, L., Hyönä, J., & Calvo, M. G. (2006). Eye movement assessment of selective attentional capture by emotional pictures. *Emotion, 6*(2), 257–268. https://doi.org/10.1037/1528-3542.6.2.257

Pollux, P. M. J., Hall, S., & Guo, K. (2014). Facial expression training optimises viewing strategy in children and adults. *PloS One, 9*(8), Article e105418. https://doi.org/10.1371/journal.pone.0105418

Raven, J., Raven, J. C., & Court, J. H. (1998). Section 1: General overview. *Manual for Raven's progressive matrices and vocabulary scales*. Oxford, UK: Oxford Psychologists Press. San Antonio, TX: The Psychological Corporation.

Rennels, J. L., & Cummings, A. J. (2013). Sex differences in facial scanning: Similarities and dissimilarities between infants and adults. *International Journal of Behavioral Development, 37*(2), 111–117. https://doi.org/10.1177/0165025412472411

Schneidt, A., Jusyte, A., & Schönenberg, M. (2019). Interpretation of ambiguous facial affect in adults with attention-deficit/hyperactivity disorder. *European Archives of Psychiatry and Clinical Neuroscience, 269*(6), 657–666. https://doi.org/10.1007/s00406-018-0879-1

Schönenberg, M., & Jusyte, A. (2014). Investigation of the hostile attribution bias toward ambiguous facial cues in antisocial violent offenders. *European Archives of Psychiatry and Clinical Neuroscience, 264*(1), 61–69. https://doi.org/10.1007/s00406-013-0440-1

Schönenberg, M., Christian, S., Gaußer, A. K., Mayer, S. V., Hautzinger, M., & Jusyte, A. (2014). Addressing perceptual insensitivity to facial affect in violent offenders: First evidence for the efficacy of a novel implicit training approach. *Psychological Medicine, 44*(5), 1043–1052. https://doi.org/10.1017/S0033291713001517

Schurgin, M. W., Nelson, J., Iida, S., Ohira, H., Chiao, J. Y., & Franconeri, S. L. (2014). Eye movements during emotion recognition in faces. *Journal of Vision, 14*(13), 11–16. https://doi.org/10.1167/14.13.14, 14.

Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and decoding facial expressions. *Psychological Science, 16*(3), 184–189. https://doi.org/10.1111/j.0956-7976.2005.00801.x

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition* (Vol. 142). Cambridge, MA: Harvard University Press.

Sullivan, S., Campbell, A., Hutton, S. B., & Ruffman, T. (2017). What's good for the goose is not good for the gander: Age and gender differences in scanning emotion faces. *The Journals of Gerontology: Series B, 72*(3), 441–447. https://doi.org/10.1093/geronb/gbv033

Suter, J. M., Byrne, M. K., Byrne, S., Howells, K., & Day, A. (2002). Anger in prisoners: Women are different from men. *Personality and Individual Differences, 32*(6), 1087–1100. https://doi.org/10.1016/S0191-8869(01)00105-2

The International Cognitive Ability Resource Team (2014). https://icar-project.com/.

Thompson, A. E., & Voyer, D. (2014). Sex differences in the ability to recognise non-verbal displays of emotion: A meta-analysis. *Cognition & Emotion, 28*(7), 1164–1195. https://doi.org/10.1080/02699931.2013.875889

Tronick, E. Z. (1989). Emotions and emotional communication in infants. *American Psychologist, 44*(2), 112–119. https://doi.org/10.1037/0003-066X.44.2.112

Vaidya, A. R., Jin, C., & Fellows, L. K. (2014). Eye spy: The predictive value of fixation patterns in detecting subtle and extreme emotions from faces. *Cognition, 133*(2), 443–456. https://doi.org/10.1016/j.cognition.2014.07.004

van Zonneveld, L., Platje, E., de Sonneville, L., van Goozen, S., & Swaab, H. (2017). Affective empathy, cognitive empathy and social attention in children at high risk of criminal behaviour. *Journal of Child Psychology and Psychiatry, 58*(8), 913–921. https://doi.org/10.1111/jcpp.12724

Waller, R., & Hyde, L. W. (2018). Callous-unemotional behaviors in early childhood: The development of empathy and prosociality gone awry. *Current Opinion in Psychology, 20*, 11–16. https://doi.org/10.1016/j.copsyc.2017.07.037

Wells, L. J., Gillespie, S. M., & Rotshtein, P. (2016). Identification of emotional facial expressions: Effects of expression, intensity, and sex on eye gaze. *PloS One, 11*(12), Article e0168307. https://doi.org/10.1371/journal.pone.0168307

Whalen, P. J., Kagan, J., Cook, R. G., Davis, F. C., Kim, H., Polis, S., & Johnstone, T. (2004). Human amygdala responsivity to masked fearful eye whites. *Science, 306*(5704), 2061. https://doi.org/10.1126/science.1103617

Wright, E. M. (2009). The measurement of psychopathy: Dimensional and taxometric approaches. *International Journal of Offender Therapy and Comparative Criminology, 53*(4), 464–481. https://doi.org/10.1177/0306624x08319416

Received: 5 May 2021 | Revised: 1 October 2021 | Accepted: 30 December 2021

DOI: 10.1002/ab.22018

RESEARCH ARTICLE

AGGRESSIVE BEHAVIOR | WILEY

Reduced attention toward faces, intentionality and blame ascription in violent offenders and community-based adults: Evidence from an eye-tracking study

Anna Maria Zajenkowska[1] | Marta Bodecka[1] | Ewa Duda[2] | Claire Lawrence[3]

[1]Institute of Psychology, Maria Grzegorzewska University, Warszawa, Poland

[2]Institute of Education, Maria Grzegorzewska University, Warszawa, Poland

[3]School of Psychology, University of Nottingham, Nottingham, UK

Correspondence

Anna Maria Zajenkowska, Institute of Psychology, Maria Grzegorzewska University, Warszawa, Poland.
Email: zajenkowska@gmail.com

Funding information

Narodowe Centrum Nauki, Grant/Award Number: UMO-2017/26/D/HS6/00258

Abstract

People typically have a strong bias in attention toward faces to help them understand social interactions. Nonetheless some people, like incarcerated offenders and psychopaths, exhibit deficits in "face reading," which may impair their interpretations, especially in case of attribution allocation in harmful events. In these cases, the ascription of intentionality is key in understanding the allocation of blame and structuring social information processing. Consequently, in the current study, in addition to typically studied intentionality and blame ascription levels (subfactors of hostile attributions), we also propose a new indicator of hostile attributions: intentionality/blame isomorphism, indicating reduced differentiation between those two factors. Violent prison inmates ($N = 63$) and community-based adults without previous history of incarceration ($N = 63$) took part in an eye-tracking study. In line with our hypotheses, offenders exhibited reduced attention orienting to faces as well as greater intentionality/blame isomorphism. In the case of both groups, people looked longer at the faces of the harm doer compared with the harm receiver. Additionally, greater intentionality/blame isomorphism predicted reduced attention to faces; however, when group status was included in the model, it became the only significant predictor of the attention to faces. Future studies should examine the origins of these gaze and attribution patterns and investigate consequences related to social perception and interactions of people prone to violence.

KEYWORDS

attention, eye-tracking, faces, hostile attributions, offenders

1 | INTRODUCTION

Criminal thinking is a subject that stimulates great scientific interest (Walters, 2020). People who have a tendency to use violence quite often use techniques of neutralization (denial of injury, e.g., "no one was hurt") (Maruna & Copes, 2005; Sykes & Matza, 1957), lack adequately reflective appraisals, such as perspective taking (Bateman & Fonagy, 2008; Carver & Scheier, 1981; Reckless, 1961) and often present hostile attributions, that is, a tendency to believe that the actor did something intentionally with hostile purpose, and is to be blamed for that (Combs et al., 2007; Dodge, 2006; Tuente et al., 2019). Interpersonal perception ranging from identifying momentary emotional facial expressions to the interpretation of complex social interactions (Bernieri, 2001) is linked to interrelated and overlapping concepts of mentalization, empathy, theory of mind, and social cognition (Gadassi & Rafaeli, 2015). Patterns in social cognition can also provide a linkage between juvenile delinquency and criminality in adulthood (Walters, 2020). Therefore, more in-depth understanding of interpersonal perception in violent individuals can examine whether

criminal behavior, particularly violent behavior, is in part, a function of social cognition (Walters, 2020).

Social cognition starts with the encoding of social cues (see Social Information Processing [SIP] model; Dodge, 2006; Tuente et al., 2019). Violent individuals display a number of perceptual distortions, which might bias the interpretation of what they see and experience. For example, impaired identification of facial affect and interpretational biases, such as hostile attributions, are often linked to aggressive and violent behaviors, and especially so among violent offenders (Gehrer et al., 2021; Schönenberg et al., 2013).

2 | ATTENTIONAL DISTORTION AND VIOLENCE

There are a number of studies that might lead to the conclusion that incarcerated male offenders display deficits in emotion recognition. For example, male incarcerated psychopaths ($n = 34$) compared with nonpsychopaths ($n = 33$) exhibit deficits in nonverbal emotional processing in case of facial affect recognition of disgust (Kosson et al., 2002). Further, male offenders ($n = 49$) with dissocial (antisocial) personality disorder as compared with matched community controls ($n = 49$) display deficits in the recognition of sad facial affect (Dolan & Fullam, 2006). Differences in the recognition of facial affect might be also related to the level of intensity of the presented emotions. For instance, aggressive male individuals with antisocial personality disorder (ASPD, $n = 32$) required significantly higher levels of emotional intensity to identify angry facial expression correctly as compared with control participants ($n = 32$), but such differences are not present in case of recognition of fearful and happy expressions (Schönenberg et al., 2013). Possibly, such inconsistencies in the type of emotion recognized might be related to the studies' methodological differences, such as for instance using black and white versus colored images (Montagne et al., 2005) or age of the respondents (adult vs. juvenile offenders) (Schönenberg et al., 2013).

It is also plausible that apart from differences in the categorization of facial emotional expressions, there is some more specific difference in relation to orientation to the eye region and eye contact. Previous work has demonstrated that incarcerated male offenders who are higher in affective psychopathy ($n = 30$) display reduce eye contact when both listening or talking to another person (Gehrer et al., 2020). Moreover, both female and male offenders, compared with controls (82 incarcerated offenders, female: $n = 36$; male: $n = 46$ and 91 controls, female: $n = 36$; male: $n = 55$), display a reduced frequency of initial attention shifts to the eyes, but not in overall attention to the eye region (i.e., relative dwell time) (Gehrer et al., 2021). While quite a lot of work has been done examining the differences in eye contact and orientation to the eye region between offenders and nonoffenders, it is not always possible for people viewing a scene to get a clear "front face-on" view of the people they are looking at. Often, people view others from a variety of perspectives and still have to work out what is happening and who is potentially to blame for any harm caused. It is clear, that by processing information from the face in general, humans make sense of others, and this ability is acquired early in childhood (Haan et al., 2002). Among prison inmate populations, personality disorders related to nonsecure attachments styles are relatively common (Fazel & Danesh, 2002), it is plausible that the early interactions with a prime caregiver were less than ideal (Bodecka et al., 2021; Ogilvie et al., 2014). Indeed, Bateman and Fonagy (2008) demonstrated that oppositional defiant disorder and callous-unemotional behaviors in older children correlate with deficits in social orienting like gazing toward caregivers during dyadic interactions (Wagner et al., 2016). Also reduced face preference among infants as young as 5 weeks old predicts higher callous-unemotional behaviors at 2.5 years (Bedford et al., 2014). It appears that the ability to recognize faces as an important "carrier" of emotional information and nonverbal affective cues is a condition sine qua non for the development and maintenance of stable interpersonal relations (see, e.g., still-face paradigm, Mesman et al., 2009; Tronick et al., 1978). As a result, the current study examines attention to protagonists' faces as the unit of interest.

3 | HOSTILE ATTRIBUTIONS

Impairments in the encoding of social cues are linked to hostile attributions (HA; Wilkowski et al., 2007). Moreover, hostile attributions may also be linked to rigidity in moral judgments associated with a reduced acknowledgment of mitigating information (e.g. Malle et al., 2014). In traditional studies on HA, a person is presented with social situations where another person either harms or hinders them (self-referent position) or harms or hinders another actor (other-referent position). They are then asked *why* this person acted in this manner. Factor analytic studies related to investigating HA with vignette-based questionnaires (e.g., AIHQ, Combs et al., 2007) have given some inconclusive results as to how HA can be understood. Some studies report that the items load on a single factor representing a unitary hostile attributional style (e.g., Buck et al., 2017; Mancuso et al., 2011), but other have found that hostile attributions are a function of the different types of relationships described in each vignette (Zajenkowska et al., 2018, Zajenkowska, Prusik, et al., 2021), such as for example friends or strangers. This differentiation of attributions indicates that social contextual factors strongly affect the tendency to make hostile attributions and that the who the actor is should be included and examined in studies (A. Jahoda et al., 2006).

A further component in the examination of hostile attribution is the level of ascribed *intentionality* as an indicator of hostile attributions (Tremblay & Belchevski, 2004; Wilkowski et al., 2007), the extent to which the harm doer should be *blamed* for their action and if the harm doer has angered them (e.g., Combs et al., 2007). While intentionality and blame attributions might occur simultaneously, they are distinct, at least in their relation to anger. Intentionality, itself, is not directly related to anger. It is, however, strongly related to blame, and blame is, in turn, strongly related to anger (Zajenkowska, Prusik, et al., 2021). As a result, intentionality

and blame attributions might plausibly form a hierarchical structure, one leading to another (from intentionality to blame and then to anger) (Zajenkowska, Prusik, et al., 2021). Intentionality ascription influences the attribution of blame in two ways: it not only amplifies perceived blame but also structures the processing of information (Malle et al., 2014). Once an act is perceived as intentional, blame can be attributed depending on the context and any mitigating or amplifying factors. In this sense, the perceiver can search for justification to ascribe either minimal blame or maximal blame (Malle et al., 2014). That is why, in addition to examining the more typically studied intentionality and blame ascription levels and their relation to each other (subfactors of hostile attributions), we also propose a new indicator of hostile attributions among adults defined as intentionality/blame isomorphism: the extent of overlap/correlation between attributions of intentionality and blame. We propose that greater overlap could be associated with a more automatic and inflexible ascription of blame, once intentionality has been judged, potentially because the processing of wider contextual cues and mitigating information is reduced. Increased isomorphism could also indicate that an individual experiences less differentiation between intentionality and blame, meaning that the likelihood of aggressive reactions increases. To give a concrete example, Bateman and Fonagy (2008) describe a woman waiting for her date to arrive. When he does not show up, she ascribes intentionality and gets very angry and wants to break up. She also cuts herself. Even though she later finds out he was late due to traffic problems, she does not change the interpretation of this situation, also does not consider it as mitigating or reducing the blame. Therefore, for her, the intentionality of his act (not showing up for the date) equals blame ascribed to him, and this, in turn, leads to an angry and (self) aggressive reaction. While it may be the case that when new information is acquired, people update initial blame judgments (Monroe & Malle, 2019), this requires some flexibility of judgment. If a persona automatically judges the intention of others as hostile, this may make them less open to new mitigating information and therefore impact their cognitive processing of subsequent data (Alicke et al., 2011; Monroe & Malle, 2017). This is typical for people who operate with a more teleological interpretational system, which represents others' actions in terms of their concrete and visible outcomes (Bateman & Fonagy, 2008). In such a case, there is no need to infer intentions or contextual variables, potentially leading to a higher correlation between intentionality and blame. This is in line with studies suggesting that attributional processes related to mental state inferences are biased and some may see that the other possesses the hostile intent because they themselves have a desire to blame the other person (Ames & Fiske, 2013; Mazzocco et al., 2004).

Studies investigating encoding patterns of violent individuals have mainly focused on the attention of those high in psychopathic traits to faces and eyes in particular (e.g., Blair et al., 2004; Deeley et al., 2006; Gehrer et al., 2019; Kosson et al., 2002; Marsh & Blair, 2008), with the common finding that psychopaths have deficits in recognizing and processing emotional cues (Edalati et al., 2016). Moreover, compared with nonpsychopathic offenders, incarcerated male psychopaths exhibit significantly reduced attention orienting toward the eyes, as indicated by absolute dwell time (Gehrer et al., 2019). There is also a large number of studies that have analyzed the more specific encoding of emotional expressions. Findings typically show that antisocial behavior is linked to impairments in recognizing fearful, sad, and surprised expressions. A meta-analytic review conducted by Marsh and Blair (2008) showed no reliable impairments in recognizing happiness, anger, or disgust expressions and deficits for recognizing fear were significantly greater than deficits for any of the other expressions. Therefore, at this stage there is some agreement that antisocial behavior, mainly in case of male participants, is linked to deficits in facial affect recognition and reduced eye contact. Gehrer et al. (2021) stress that her study was the first to investigate violent female offenders' gaze patterns and that, thus far, investigating the gaze patterns of violent offenders has almost exclusively included male participants. This is important in light of the fact that some posit a general effect of gender on the perception of facial expressions (Campbell et al., 2002; Montagne et al., 2005). For example, in a study among students ($N = 68$, women $n = 40$), men were less accurate in labeling facial expressions (in particular, sadness and surprise), and also less sensitive in labeling facial expressions (anger and disgust) (Montagne et al., 2005). Studies including larger groups ($N = 240$) led to conclusion that women were faster than men in accurately identifying facial emotions, but at the same time overall sex differences in accuracy were nonexistent (Rahman et al., 2004). Based on the aforementioned results, it is clear that examining both men and women in the investigation of social cognition of intention and blame, as well as the examination of offenders versus nonoffenders will be pertinent to further our understanding of how offenders process social information involving harmful interactions.

It is also important to notice that in studies on facial affect recognition, the stimuli used is typically a "forward facing" face, rather than stimuli depicting the whole person (Duchowski et al., 2019). Nonetheless, more recently, the role of additional characteristics of the person being perceived has gained more attention. For example, people tend to avoid eye contact with those who intentionally harm others, and are more punitive toward the harm doer when they are cued to attend to faces of norm violators (Fincher, 2019). Still, the extent to which these effects vary as a function of the perceiver has not been examined yet. Considering that looking at somebody's face humanizes the perceived person (Fincher, 2019) and that noticing a face is a first step to recognize emotional expression, without which eliciting empathy and later inhibiting aggressive behavior would not be possible (Marsh & Blair, 2008), investigating attention to faces of both harm doer and receiver in social encounters among people who committed violent crimes is important and, to our knowledge, has not been done yet.

4 | CURRENT STUDY

The main focus in the current project is the extent to which offenders versus nonoffenders look toward the face region as an initial primary region of curiosity to understand a scene where harm is caused to one person by another, and which of the protagonists is the main focus of interest—the person who is harmed or the person doing the harming. Therefore, we investigated group differences between male and female violent offenders and community-based adults in their general attention orienting to the faces (dwell time[1] during the stimulus presentation) and in the attention orientation on the harm doer versus harm receiver faces in ambiguous visual scenes. Interestingly, inmates tend to declare lower hostile attribution in the case of social perception of provocations (Zajenkowska et al., 2013, 2018; Zajenkowska, Rogoza, et al., 2021). As a result, we investigated group differences not only in ascribed intentionality, and blame but also in the newly proposed intentionality/blame isomorphism. Finally, we wanted to better understand the cognitive mechanism behind intentionality/blame isomorphism and how it is linked to attention to faces.

First, we hypothesize that in a gender-balanced group of community adults without prior history of incarceration and offenders, the latter will present reduced attention toward the faces of both the harm doers and receivers (H1). Further, because from an evolutionary perspective precise recognition of a threatful, provocative or aggressive action is an adaptive mechanism (Dodge, 2006), both offenders and nonoffenders are expected to attend longer (longer dwell time) to the harm doers' faces as compared with harm receivers' faces (H2). Nonetheless, group differences could emerge especially in the case of the harm receivers' faces due to lower empathy in violent offenders (Jolliffe & Farrington, 2004). Consequently, we expected violent offenders to attend less than nonoffenders to harm receivers' faces as compared with harm doers (H3). Previous findings, mainly among violent young offenders, show links between hostile attributions and aggression (Dodge, 2006). However, at the same time, inmates declare lower levels of sensitivity to provocation or hostile attributions (Zajenkowska et al., 2014; Zajenkowska, Prusik, et al., 2018; 2021) compared with nonoffenders. This paradox might suggest that the mechanism leading to aggression and violence is not linked simply to the intensity of declared HA. Therefore, we postulate that intentionality/blame isomorphism will be higher in violent offenders compared with nonoffenders (H4). Finally, because higher intentionality/blame isomorphism might indicate individuals' lack of judgment flexibility and consideration or exploration of situational cues, we wanted to explore if intentionality/blame isomorphism remains linked to reduced attention to faces when controlling for the effect of group: offenders versus community-based adults (H5).

The current study aims to add to the knowledge about techniques of neutralization of harm (Walters, 2020), namely (1) avoiding attention to informative regions of an actor—in this case, the face of harm doer or harm receiver, (2) and reduced reflective reappraisal exemplified by high correlation of intentionality and blame: intentionality/blame isomorphism.

5 | METHOD

5.1 | Participants

Two groups of participants took part in the study: violent offenders ($N = 63$, $M_{age} = 38.4$ years, SD = 9.4 years, range = 22–60, women $n = 31$); and community-based adults with no previous history of incarceration recruited to best match the offender group ($N = 63$, $M_{age} = 33.1$ years, SD = 8.9 years, range = 23–57, women $n = 31$). A priori calculation of statistical power G*Power (Faul et al., 2009) for regression with four predictors indicated that at least 118 participants were needed to detect a medium size effect (Cohen's $d = 0.65$) with power $(1 - \beta) = .80$ and $\alpha = .05$. In the case of three participants, information about age was missing; in the case of one participant information about education level was missing. As the number of participants in the noninmates group was initially twice that of the inmate group, participants in the community-based adults' group were chosen to match at a group level in terms of age, sex, and education; therefore, the youngest participants and those with the highest levels of education were excluded from the analyses.

In case of violent offenders, we informed the prison authorities that the individuals eligible to participate would only be those with "any criminal charge for a violent offense against persons—e.g., assault, assault causing bodily harm, wounding, attempted homicide, homicide, kidnapping, forcible confinement, armed robbery, and all 'hands-on' sexual offenses" (Harris et al., 2002, p. 383). In the case of the nonincarcerated group, we asked participants about their previous imprisonments and only those with no such previous imprisonment were included in the group. Demographic data of all four groups (offenders and community-based adults; female and male participants) are displayed in Table 1. Incarcerated offenders serving prison sentences for violent crimes (e.g., murder, attempted murder, physical assault) from two correctional facilities were invited to participate in the study by corrections officers. Those who agreed to participate met individually with the researcher in a designated quiet room within the prisons. The community sample was recruited through social media or invited directly by the researchers. Community participants met individually with the researcher in a designated quiet room at the university. Participants were first prescreened to ensure they comply with inclusion criteria: no psychotic disorders and no serious vision problems (e.g., for eye-tracking methodology).

The study was approved by the (anonymous) university's ethics committee and conducted in accordance with the Declaration of Helsinki.

6 | PROCEDURE AND MEASURES

The study was part of a larger project; therefore, the research procedure consisted of several elements: first, the participants took part in the task of observing visual scenes, then they performed a second (data reported elsewhere) computer-based task, and finally they

TABLE 1 Education and age in offenders and community-based adults (nonoffenders)

	Female offenders (*n* = 31)	Female nonoffenders (*n* = 31)	Male offenders (*n* = 32)	Male nonoffenders (*n* = 32)
Age (years)	37.00 (7.57)	31.71 (10.09)	39.62 (10.75)	34.41 (7.50)
Fluid intelligence	2.81 (2.04)	5.26 (2.62)	2.69 (2.12)	5.34 (2.58)
Education level				
Primary	48.4%	0%	40.6%	6.3%
Secondary	32.3%	16.1%	40.6%	37.4%
High school	16.1%	51.6%	12.6%	34.4%
Higher	3.2%	32.3%	3.1%	21.9%
Not reported	0%	0%	3.1%	0%

Note: The data represented in the table refer to means and standard deviations (in parentheses) for age and to relative frequencies for education level.

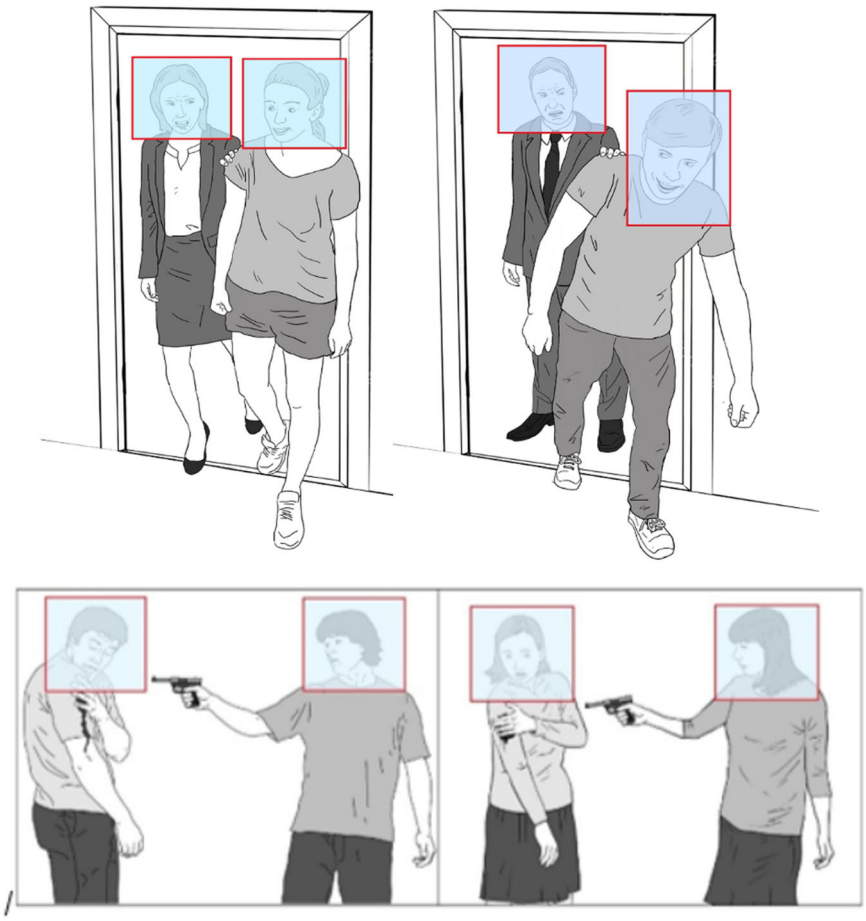


FIGURE 1 Example of an illustration of the definition of the AOI. AOI, area of interest

completed a series of personality questionnaires (data reported elsewhere).

Task-observing visual scenes: participants were asked to complete the task while their eye movements were recorded by an eye tracker. The participants were asked to spontaneously look at each scene. There was no fixation cross. The stimuli[2] consisted of pictures (2126 × 1594 pixels) depicting scenes with two actors—a harm doer and a harm receiver. All either scenes depicted peers interacting with each other, or subordinates and authority figures (see Figure 1 for examples). For scenes with peers, each scene depicted either two men or two women to avoid associations with heterosexual intimate partner violence. Graphics were used to create new scenes depicting women, which were

based on the scenes depicting men used in the study by Wilkowski et al. (2007). To ensure that the facial expression and gestures were typical for females, two women posed and were sketched for each of the scenes. In the complete corpus of stimuli, scenes were either clearly hostile, nonhostile or ambiguous and have been validated accordingly (Wilkowski et al., 2007; Zajenkowska & Rajchert, 2020). For the purpose of the current study, we focused solely on the 54 ambiguous scenes where the intentions of the actor were unclear and could be interpreted as intentional to various degrees (Wilkowski et al., 2007, Zajenkowska & Rajchert, 2020). Each scene was displayed on a computer screen. All scenes were black and white monochrome, the shade of hair of both the harm doer and receiver was randomized (but not equally distributed) across trials, as well as the position (left/right) that the harm doer was placed. On a subsequent screen, participants made, first, intentionality and then blame judgments, for example, "Please rate to what extent the depicted harm was intentional" on a Likert scale ranging from 1 (not intended at all) to 9 (intended) and to what extent you would blame the person for that (1 not at all to 9 very much) (see Zajenkowska & Rajchert, 2020).

Intentionality and blame isomorphism was calculated using the R Project for Statistical Computing. First, to create a total score of intentionality and blame for each type of scene, we averaged the responses for all images ($\alpha_{\text{intentionality}} = .91$; $\alpha_{\text{blame}} = .93$). Then the average correlation between intentionality and blame was calculated for each participant.

Fluid intelligence[3] was assessed with the 11 Matrix Reasoning items from the International Cognitive Ability Resource (ICAR; The International Cognitive Ability Resource Team, 2014). The Matrix Reasoning items contain stimuli similar to Raven's Progressive Matrices (Raven & John Hugh Court., 1998), including arrays with geometric shapes with one of the nine shapes missing and six geometric shapes displayed as response alternatives, from which participants choose which one best completes the arrays.

6.1 | Eye-tracking measures

Eye movements were measured using a Tobii Pro X3-120 remote eye tracker with a sampling rate of 120 Hz, and the binocular accuracy 0.5^{0}/precision 0.25^{0} [4]. Data recording was done using a Dell laptop, stimuli were presented on a 17″ computer screen (1920 × 1080 pixels resolution) for 6–9 s, as the varying presentation times reduce the top-down-driven influences on the gaze patterns because they diminish predictability of stimulus onset and offset. Participants viewed the presented scenes from a distance of ca. 60–70 cm from the screen. The study was preceded by a 9-point calibration, repeated until a satisfactory result was obtained.[5]

6.2 | Data analysis

The analysis of the gaze patterns of the ambiguous scenes was conducted using the iMotions Attention Tool software (version 7.2; iMotions A/S). Faces of both the harm doers and harm receivers were defined as the area of interest (AOI; a frame including whole faces, see Figure 1) and using iMotion software, which allows also for obtaining statistical metrics for each AOI (e.g., fixation duration, saccades). The AOI size differed as a function of the position of the actor in the scene—as the scenes presented two interacting people and included a visual perspective, each face could have been closer or further; therefore, the shape and size of the AOI were slightly different (square or rectangle, side length from ca. 2 to ca. 6 cm, ca. 2–5° visual angle); also the faces could have been presented as forward facing or in profile (see Figure 1). The faces' AOIs positions mirrored those in plausible life situations. Moreover, the AOI size depended on the size of the face and was defined based on the recommendation by Holmqvist et al. (2011) that indicated a buffer space (margin) to ensure inclusion of all fixations belonging to a given object. Because in our study only a few scenes included overlapping objects (e.g., faces and hand), in such situations analyses benefit from larger AOI sizes (>0.5° visual angle margins) (Jayawardena et al., 2020; Orquin et al., 2016; Duchowski et al., 2019).

Dwell time was defined as the percentage of total available time spent in AOI calculated separately for each AOI. To make sure that respondents saw the face and not just saccaded through it, dwell time was calculated using fixation-based metrics. Fixations were qualified using the Identification by Velocity Threshold (IVT) algorithm (Salvucci & Goldberg, 2000). Dwell time indicated the time spent in AOI, based on total duration of all respondents' fixations (excluding data points between fixations) and therefore was treated as a measure of general attention to faces.

7 | RESULTS

Correlation coefficients between encoding indices (dwell times) and interpretation indices are presented in Table 2. Intentionality/blame isomorphism was negatively related to the attention to faces, but also to fluid intelligence.

Group differences in age were investigated by 2 (group) × 2 (gender) ANOVAs. Age differed significantly between groups, which was indicated by a significant main effect of group, $F(1, 119) = 10.15$, $p < .005$, $\eta_p^2 = 00.079$ with offenders being older. The main effect of gender, $F(1, 119) = 2.60$, $p = .11$ and interaction of group and gender, $F(1, 119) < 0.01$, $p = .98$ remained nonsignificant. Fluid intelligence (ICAR) differed significantly between groups, which was indicated by a significant main effect of group, $F(1, 122) = 37.02$, $p < .001$, $\eta_p^2 = 0.233$ with nonoffenders with higher ICAR. The main effect of gender, $F(1, 122) < 0.01$, $p = .97$ and interaction of group and gender, $F(1, 122) = 0.06$, $p = .81$ remained nonsignificant. Furthermore, education level was generally higher in nonoffenders participants compared with offenders (see Table 1). Because education, intelligence, and age are related to cognitive processes, these were controlled for in the analysis.

TABLE 2 Pearson's correlations between hostile attributions subfactors, attention to faces, and fluid intelligence (ICAR)

		1	2	3	4	5	6	7
1	I/B isomorphism	–	−0.032	0.084	−0.221*	−0.188*	−0.247**	−0.190*
2	Intentionality	−0.032	–	0.772**	0.036	0.012	0.062	0.090
3	Blame	0.084	0.772**	–	0.089	0.080	0.093	0.070
4	Dwell_time_overall	−0.221*	0.036	0.089	–	0.982**	0.977**	0.277**
5	Dwell_time_harm_doer	−0.188*	0.012	0.080	0.982**	–	0.919**	0.268**
6	Dwell_time_harm_receiver	−0.247**	0.062	0.093	0.977**	0.919**	–	0.272**
7	ICAR	−0.190*	0.090	0.070	0.277*	0.268*	0.272*	–

*Correlation is significant at the 0.05 level (two-tailed).

**Correlation is significant at the 0.01 level (two-tailed).

7.1 | Attention to faces

To investigate dwell time on faces between offenders and nonoffenders, we conducted a hierarchical regression. Covariates (age, level of education, and ICAR) were entered in the first block and group (offender vs. nonoffender) were entered in the second block. The results revealed a significant effect of group, indicating a longer duration of dwell time for faces for nonoffenders than offenders, $R^2 = .11$, $F(4, 117) = 4.90$, $p < .005$. Apart from group, which was significantly associated with the dwell time ($\beta = .24$, $t = 2.09$, $p = .039$) other variables (excluded from the model) remained nonsignificant predictors of the attention to faces, age ($\beta = -.14$, $t = -1.56$, $p = .12$), level of education ($\beta = -.07$, $t = -0.64$, $p = .52$), ICAR ($\beta = .17$, $t = 1.65$, $p = .10$), indicating that difference in terms of dwell time is fully explained by group. A medium effect size (Cohen's $d = 0.72$), Power ($1 - \beta$ err prob) = 0.90 was validated by post hoc G*Power analysis.

To determine whether offenders and nonoffenders showed different levels of attention to faces of the harm doer versus harm receiver, a mixed-measures analyses of variance was conducted. The analyses revealed that dwell time varied by group and type of actor. Mixed-measures analyses revealed a significant main between-group effect of group, $F(1, 124) = 14.84$, $p < .001$, medium effect size (Cohen's $d = 0.69$), Power ($1 - \beta$ err prob) = 0.99, indicating longer dwell time for faces among nonoffenders ($M = 18.85\%$, SE = 0.90%) compared with the offenders ($M = 13.92\%$, SE = 0.90%). The analyses revealed a significant main within-group effect of the actor, $F(1, 124) = 39.70$, $p < .001$, large effect size (Cohen's $d = 1.13$), Power ($1 - \beta$ err prob) = 0.99, indicating longer dwell time for faces of the harm doers ($M = 17.33\%$, SE = 0.72%) compared with the harm receivers ($M = 15.44\%$, SE = 0.59%). The interaction between group and type of actor was not significant, $F(1,124) = 0.28$, $p = .59$.

7.2 | Intentionality, blame ascription as well as intentionality/blame isomorphism

To explore the effect of group type on intentionality and blame ascription, an independent t-test was conducted. The intentionality ascription was higher for nonoffenders ($M = 6.01$, SE = 0.11) compared with offenders ($M = 5.59$, SE = 0.18), $t(124) = -2.02$, $p = .023$ (one-tailed), $d_{Cohen} = 0.36$, 95% CI [−0.84, −0.01]. The blame ascription was also higher for nonoffenders ($M = 5.54$, SE = 0.14) compared with offenders ($M = 4.99$, SE = 0.19), $t(124) = -2.33$, $p = .011$ (one-tailed), $d_{Cohen} = 0.40$, 95% CI [−1.02, −0.08].

To test the hypothesis that offenders and nonoffenders differ in terms of intentionality/blame, we conducted a hierarchical regression. Covariates (age, level of education, and ICAR) were entered in the first block and group (offender vs. nonoffender) were entered in the second block. The model was significant, $R^2 = .12$, $F(4, 117) = 5.27$, $p < .005$. Results revealed that group was a significant predictor ($\beta = -.31$, $t = -2.78$, $p < .01$), indicating higher intentionality/blame isomorphism for offenders than nonoffenders, and age ($\beta = .21$, $t = 2.29$, $p < .05$), suggesting that intentionality/blame isomorphism increases with age. Other variables remained nonsignificant: level of education ($\beta = .17$, $t = 1.56$, $p = .12$), ICAR ($\beta = -.016$, $t = -0.63$, $p = .53$). A medium effect size (Cohen's $d = 0.75$), Power ($1 - \beta$ err prob) = 0.93 was validated by post hoc G*Power analysis.

Finally, to test the hypothesis that intentionality/blame isomorphism is related to dwell time (on all faces), a regression analysis was conducted. The analyses revealed a significant effect of intentionality/blame isomorphism, $R^2 = .04$, $F(1, 124) = 6.38$, $p < .05$, small effect size (Cohen's $d = 0.41$), Power ($1 - \beta$ err prob) = 0.63. Specifically, with increased isomorphism, the dwell time on faces reduced, $\beta = -.22$, $t(124) = -2.52$, $p < .05$, 95% CI [−9.69, −1.17]. In the second step, group was included and the analyses revealed a significant R^2 change, $R^2 = .11$, $F(2, 123) = 8.77$, $p < .001$, medium effect size (Cohen's $d = 0.70$), Power ($1 - \beta$ err prob) = 0.95. At this point, group was the only significant predictor of the dwell time on faces, $\beta = .29$, $p < .005$, 95% CI [1.70, 6.93], and isomorphism became nonsignificant, $p = .14$.[6]

8 | DISCUSSION

In the present study, initially, we investigated group differences between male and female violent offenders and community adults in general attention orienting to the faces, but also to the harm doers' and harm receivers' faces in ambiguous scenes by examining dwell

time during stimulus presentation. As predicted, compared with community adults, offenders presented reduced attention orienting to all faces. This finding mirrors other work demonstrating emotional processing deficits in violent offenders, but the current study is unique in that it looks at both male and female participants in both offender and nonoffender groups (Gehrer et al., 2019). Also, in line with the hypothesis, all participants looked more at the faces of harm doers compared with those who were harmed in the scenes. This finding supports evolutionary explanations for attention to threatening faces. That is, increased attention to faces is more informative in directing our attention to where the threat is, and therefore likely to help us avoid harm (Fox et al., 2000). For the first time, this study has also shown this effect when participants can split attention simultaneously between the person causing harm and the person being harmed. This is a significant step and allows us to compare which of these two individuals capture attention. Our finding that the harm doer's face captures more attention than the harm receivers' face is consistent with the findings of Green et al. (2003) who showed that angry faces (indicating the potential for doing harm to the perceiver) were processed more efficiently and comprehensively than fearful or sad faces (harm already done, or about to be done, to someone else). In future studies, in addition to general attention to faces, the impact of variable time of presentation of each scene on proportional reduction/increase of dwell time could be tested. This would allow an examination of whether aggressive individuals lose interest and/or attention focus across the presentation of stimuli or whether they pay less attention to faces regardless of the duration of the presentation.

Contrary to our expectation, there were no differences in the extent to which community adults and offenders attended to the faces of the harm receivers as compared with the harm doers. Past findings that show consistent deficits in offenders' processing of fearful or sad faces would suggest that attention to the harm receiver would have been particularly reduced in the offender groups. Previous work, however, has measured the *accuracy* of emotions and perceivers' specific eye movements across the whole face stimuli. We suggest that the simultaneous pairing of both harm doer and harm receiver is ecologically more valid, but yields a pattern of attention between the faces of the harm doer (potential threat) and the harm receiver that is more similar to that of nonoffenders. Future work should address issues of both the function of the actor in the scene and the emotional expression on the face. To conclude, on one hand, violent offenders presented lower attention toward faces in general, but the distributed pattern of looking at the harm doer and harm receiver matched that of community-based adults—longer dwell times were orientated to the harm doer.

Regarding the interpretation of the ambiguous harmful acts, inmates and noninmates did present lower levels of both ascribed intentionality and blame, which is consistent with previous findings (Zajenkowska et al., 2018, 2020). Nonetheless, as stipulated, the difference was observed in intentionality/blame isomorphism, which can be interpreted in line with the hypotheses of threshold theory (Karwowski & Gralewski, 2013) or Spearman's law of diminishing returns (Murray et al., 2013), which states that correlations between constructs are weaker for higher ability (e.g., cognitive, intellectual) levels. It is feasible that the more cognitive flexibility a person possesses, the more they are able to differentiate between constructs and consequently a more sophisticated moral judgment develops. Conversely, among those lower in cognitive or mentalizing ability, such as the inmates in the current study and elsewhere (Bateman & Fonagy, 2008), intentionality inferences lead to blaming. As a result, those concepts might be synonymous. However, in many cases, an agent's action may be intentional, but the outcome or harm is not solely within the control or cause of the individual themselves. As a result, there is less overlap between attributions of intent and blame (Malle & Bennett, 2002). Such interpretation requires differentiation between those two concepts: blame and intentionality, and from our study we can conclude that the differentiation is less common among violent offenders as indicated by higher intentionality/blame isomorphism. Moreover, higher correlations between intentionality and blame may stem from reduced ability to deal with ambiguity when both hostile and not-hostile cues coexist (e.g. current study material—visual scenes). In such cases, humans tend to look for illusory correlations and apply more automatic processing (Gilovich et al., 2002). Whether this tendency for automatic processing and cognitive heuristics such as illusory correlations is observed more in those with lower mentalizing ability needs further research.

Finally, our results revealed that although greater intentionality/blame isomorphism predicted reduced attention to faces when the group was included in the model, it became the only significant predictor of dwell time to faces. On one hand, such results are in line with the notion that certain maladaptive schemata impact the encoding patterns (Wilkowski et al., 2007). Nonetheless, there are some other factors among inmates that are linked to lesser attention to faces, possibly related to some individual characteristics common to this group. For example, socially anxious individuals, who typically are concerned about negative evaluation by others, present attentional bias away from faces expressing both positive and negative emotions (Mansel et al., 1999). It is possible that violent offenders experienced elevated social-evaluative threat, compared with community-based participants, given their penal context. Consequently, future studies examining attention to faces should also consider the context and meaning of the study for participants.

In this sense, the fact that the study was conducted in prison is a limitation of the current work. Future studies could utilize online tools to conduct such studies, possibly reducing the in-person contact with an investigator. Additionally, including more indicators of encoding could help to develop a better understanding of the bases of hostile attributions and particularly intentionality/blame isomorphism.

The current paper demonstrates the differences between offenders and nonoffenders in the social perception of scenes depicting harm. The findings suggest that differences may be based, at least in part, on the extent to which observers hold a more overlapping view of intent and blame. Further work on this intention/

blame isomorphism could yield fruitful seams of investigation in the field of hostile attributions both in forensic and nonforensic settings.

DATA AVAILABILITY STATEMENT

Data used in this study will be stored at: http://www.apsycholab.pl

ORCID

Anna Maria Zajenkowska https://orcid.org/0000-0002-9413-1428

ENDNOTES

[1] When investigating attentional biases, one can focus on a more automatic mode or more effortful, although still influenced by preexisting intrapersonal schemas (Beck & Clark, 1997; Cisler & Koster, 2010). The first one is quick, often unintentional orienting to a stimuli, which could be measured by time to first fixation or initial attention shifts and the latter could be indicated by the time spent on the stimuli, for example, dwell time (Gehrer et al., 2021).

[2] All stimuli can be found and downloaded from http://apsycholab.pl/team/

[3] We decided to control for possible effects of higher intelligence in offenders compared with nonoffenders; therefore, ICAR was one of the measures (see Gehrer et al., 2021).

[4] https://www.tobiipro.com/siteassets/tobii-pro/accuracy-and-precision-tests/tobii-pro-x3-120-accuracy-and-precision-test-report.pdf

[5] In case of few participants the calibration was less than good; that is why we checked if excluding them would impact the results. Because it did not, and also because the eye-tracking data were tied to behavioral data (e.g., intentionality assessment) we decided not to exclude those participants.

[6] Additionally, we have conducted separate analysis for both harm doers' and harm receivers' faces and results were analogous (harm doers' faces: group was the only significant predictor of the dwell time on harm doers' faces, $\beta = .25$, $p < .01$, 95% CI [1.20, 7.18], and isomorphism became nonsignificant, $p = .23$; harm receivers' faces: group was the only significant predictor of the dwell time on harm receivers' faces, $\beta = .32$, $p < .001$, 95% CI [2.01, 6.86], and isomorphism became nonsignificant, $p = .09$.

REFERENCES

Alicke, M. D., Rose, D., & Bloom, D. (2011). Causation, norm violation, and culpable control. *Journal of Philosophy*, *108*(12), 670–696.

Ames, D. L., & Fiske, S. T. (2013). Intentional harms are worse, even when they're not. *Psychological Science*, *24*(9), 1755–1762.

Bateman, A., & Fonagy, P. (2008). Comorbid antisocial and borderline personality disorders: Mentalization-based treatment. *Journal of Clinical Psychology*, *64*(2), 181–194.

Beck, A. T., & Clark, D. A. (1997). An information processing model of anxiety: Automatic and strategic processes. *Behaviour Research and Therapy*, *35*(1), 49–58.

Bedford, R., Pickles, A., Sharp, H., Wright, N., & Hill, J. (2014). Archival report reduced face preference in infancy: A developmental precursor to callous-unemotional traits? *Biological Psychiatry*, *78*, 1–7.

Bernieri, F. J. (2001). Toward a taxonomy of interpersonal sensitivity. In *Interpersonal sensitivity: Theory and measurement* (pp. 3–20).

Blair, R. J. R., Mitchell, D. G. V., Peschardt, K. S., Colledge, E., Leonard, R. A., Shine, J. H., Murray, L. K., & Perrett, D. I. (2004). Reduced sensitivity to others' fearful expressions in psychopathic individuals. *Personality and Individual Differences*, *37*(6), 1111–1122.

Bodecka, M., Grala, K., Szymczak, K., Zajenkowska, A., Rajchert, J., Kostrzewska, D., & Lawrence, C. (2021). Relationship between trait anger and rejection sensitivity: The moderating role of the group (inmates and non-inmates). *The Journal of Forensic Psychiatry & Psychology*, *32*, 1–16.

Buck, B., Iwanski, C., Healey, K. M., Green, M. F., Horan, W. P., Kern, R. S., Lee, J., Marder, S. R., Reise, S. P., & Penn, D. L. (2017). Improving measurement of attributional style in schizophrenia: A psychometric evaluation of the Ambiguous Intentions Hostility Questionnaire (AIHQ). *Journal of Psychiatric Research*, *89*, 48–54.

Campbell, R., Elgar, K., Kuntsi, J., Akers, R., Terstegge, J., Coleman, M., & Skuse, D. (2002). The classification of 'fear'from faces is associated with face recognition skill in women. *Neuropsychologia*, *40*(6), 575–584.

Carver, C. S., & Scheier, M. F. (1981). Self-consciousness and reactance. *Journal of Research in Personality*, *15*, 16–29.

Cisler, J. M., & Koster, E. H. (2010). Mechanisms of attentional biases towards threat in anxiety disorders: An integrative review. *Clinical Psychology Review*, *30*(2), 203–216.

Combs, D. R., Penn, D. L., Wicher, M., & Waldheter, E. (2007). The Ambiguous Intentions Hostility Questionnaire (AIHQ): A new measure for evaluating hostile social-cognitive biases in paranoia. *Cognitive Neuropsychiatry*, *12*(2), 128–143.

Deeley, Q., Daly, E., Surguladze, S., Tunstall, N., Mezey, G., Beer, D., Ambikapathy, A., Robertson, D., Giampietro, V., Brammer, M. J., Clarke, A., Dowsett, J., Fahy, T., Phillips, M. L., & Murphy, D. G. (2006). Facial emotion processing in criminal psychopathy: Preliminary functional magnetic resonance imaging study. *The British Journal of Psychiatry*, *189*(6), 533–539.

Dodge, K. A. (2006). Translational science in action: Hostile attributional style and the development of aggressive behavior problems. *Development and Psychopathology*, *18*(3), 791–814.

Dolan, M., & Fullam, R. (2006). Face affect recognition deficits in personality-disordered offenders: Association with psychopathy. *Psychological Medicine*, *36*(11), 1563–1569.

Duchowski, A. T., Gehrer, N. A., Schönenberg, M., & Krejtz, K. (2019, June). Art facing science: Artistic heuristics for face detection: Tracking gaze when looking at faces. In *Proceedings of the 11th ACM symposium on eye tracking research & applications* (pp. 1–5).

Edalati, H., Walsh, Z., & Kosson, D. S. (2016). Attentional bias in psychopathy: An examination of the emotional dot-probe task in male jail inmates. *International Journal of Offender Therapy and Comparative Criminology*, *60*(11), 1344–1357.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160.

Fazel, S., & Danesh, J. (2002). Serious mental disorder in 23 000 prisoners: A systematic review of 62 surveys. *The Lancet*, *359*(9306), 545–550.

Fincher, K. M. (2019). Social antecedents and perceptual consequences of how we look at others. *Journal of Experimental Psychology: General*, *148*(1), 143–157.

Fox, E., Lester, V., Russo, R., Bowles, R. J., Pichler, A., & Dutton, K. (2000). Facial expressions of emotion: Are angry faces detected more efficiently? *Cognition & Emotion*, *14*, 61–92.

Gadassi, R., & Rafaeli, E. (2015). Interpersonal perception as a mediator of the depression–interpersonal difficulties link: A review. *Personality and Individual Differences*, *87*, 1–7.

Gehrer, N. A., Duchowski, A. T., Jusyte, A., & Schönenberg, M. (2020). Eye contact during live social interaction in incarcerated psychopathic offenders. *Personality Disorders: Theory, Research, and Treatment*, *11*(6), 431–439.

Gehrer, N. A., Scheeff, J., Jusyte, A., & Schönenberg, M. (2019). Impaired attention toward the eyes in psychopathic offenders: Evidence from

an eye tracking study. *Behaviour Research and Therapy*, *118*, 121–129.

Gehrer, N. A., Zajenkowska, A., Bodecka, M., & Schönenberg, M. (2021). Attention orienting to the eyes in violent female and male offenders: An eye-tracking study. *Biological Psychology*, *163*, 108136.

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge University Press.

Green, M. J., Williams, L. M., & Davidson, D. (2003). In the face of danger: Specific viewing strategies for facial expressions of threat? Brief Report. *Cognition & Emotion*, *17*(5), 779–786.

Haan, M. D., Pascalis, O., & Johnson, M. H. (2002). Specialization of neural mechanisms underlying face recognition in human infants. *Journal of Cognitive Neuroscience*, *14*(2), 199–209.

Harris, G. T., Rice, M. E., & Cormier, C. A. (2002). Prospective replication of the Violence Risk Appraisal Guide in predicting violent recidivism among forensic patients. *Law and Human Behavior*, *26*(4), 377–394.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP.

Jahoda, A., Pert, C., & Trower, P. (2006). Frequent aggression and attribution of hostile intent in people with mild to moderate intellectual disabilities: An empirical investigation. *American Journal on Mental Retardation*, *111*(2), 90–99.

Jayawardena, G., Michalek, A., Duchowski, A., & Jayarathna, S. (2020, June). Pilot study of audiovisual speech-in-noise (sin) performance of young adults with ADHD. In *ACM symposium on eye tracking research and applications* (pp. 1–5).

Jolliffe, D., & Farrington, D. P. (2004). Empathy and offending: A systematic review and meta-analysis. *Aggression and Violent Behavior*, *9*(5), 441–476.

Karwowski, M., & Gralewski, J. (2013). Threshold hypothesis: Fact or artifact? *Thinking Skills and Creativity*, *8*, 25–33.

Kosson, D. S., Suchy, Y., Mayer, A. R., & Libby, J. (2002). Facial affect recognition in criminal psychopaths. *Emotion*, *2*(4), 398–411.

Malle, B. F., & Bennett, R. E. (2002). *People's praise and blame for intentions and actions: Implications of the folk concept of intentionality*. Institute of Cognitive and Decision Sciences, University of Oregon.

Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A theory of blame. *Psychological Inquiry*, *25*(2), 147–186.

Mancuso, F., Horan, W. P., Kern, R. S., & Green, M. F. (2011). Social cognition in psychosis: Multidimensional structure, clinical correlates, and relationship with functional outcome. *Schizophrenia Research*, *125*(2-3), 143–151.

Marsh, A. A., & Blair, R. J. R. (2008). Deficits in facial affect recognition among antisocial populations: A meta-analysis. *Neuroscience & Biobehavioral Reviews*, *32*(3), 454–465.

Maruna, S., & Copes, H. (2005). What have we learned from five decades of neutralization research? *Crime and Justice*, *32*, 221–320.

Mazzocco, P. J., Alicke, M. D., & Davis, T. L. (2004). On the robustness of outcome bias: No constraint by prior culpability. *Basic and Applied Social Psychology*, *26*(2), 131–146.

Mesman, J., van IJzendoorn, M. H., & Bakermans-Kranenburg, M. J. (2009). The many faces of the Still-Face Paradigm: A review and meta-analysis. *Developmental Review*, *29*(2), 120–162.

Monroe, A. E., & Malle, B. F. (2017). Two paths to blame: Intentionality directs moral information processing along two distinct tracks. *Journal of Experimental Psychology: General*, *146*(1), 123–133.

Monroe, A. E., & Malle, B. F. (2019). People systematically update moral judgments of blame. *Journal of Personality and Social Psychology*, *116*(2), 215–236.

Montagne, B., Kessels, R. P., Frigerio, E., de Haan, E. H., & Perrett, D. I. (2005). Sex differences in the perception of affective facial expressions: Do men really lack emotional sensitivity? *Cognitive Processing*, *6*(2), 136–141.

Murray, A. L., Dixon, H., & Johnson, W. (2013). Spearman's law of diminishing returns: A statistical artifact? *Intelligence*, *41*(5), 439–451.

Ogilvie, C. A., Newman, E., Todd, L., & Peck, D. (2014). Attachment & violent offending: A meta-analysis. *Aggression and Violent Behavior*, *19*(4), 322–339.

Orquin, J. L., Ashby, N. J., & Clarke, A. D. (2016). Areas of interest as a signal detection problem in behavioral eye-tracking research. *Journal of Behavioral Decision Making*, *29*(2–3), 103–115.

Rahman, Q., Wilson, G. D., & Abrahams, S. (2004). Sex, sexual orientation, and identification of positive and negative facial affect. *Brain and Cognition*, *54*(3), 179–185.

Raven, J. C., & John Hugh Court. (1998). *Raven's progressive matrices and vocabulary scales* (Vol. 759). Oxford Pyschologists Press.

Reckless, W. C. (1961). A new theory of delinquency and crime. *Federal Probation*, *25*(4), 42–46.

Salvucci, D. & Goldberg, J. (2000). Identifying fixations and saccades in eye-tracking protocols. In Proceedings of the 2000 symposium on eye tracking research & applications (pp. 71–78). ACM.

Schönenberg, M., Louis, K., Mayer, S., & Jusyte, A. (2013). Impaired identification of threat-related social information in male delinquents with antisocial personality disorder. *Journal of Personality Disorders*, *27*(4), 496–505.

Sykes, G. M., & Matza, D. (1957). Techniques of neutralization: A theory of delinquency. *American Sociological Review*, *22*(6), 664–670.

Tremblay, P. F., & Belchevski, M. (2004). Did the instigator intend to provoke? A key moderator in the relation between trait aggression and aggressive behavior. *Aggressive Behavior*, *30*(5), 409–424.

Tronick, E., Als, H., Adamson, L., Wise, S., & Brazelton, T. B. (1978). Infants response to entrapment between contradictory messages in face-to-face interaction. *Journal of the American Academy of Child and Adolescent Psychiatry*, *17*, 1–13.

Tuente, S. K., Bogaerts, S., & Veling, W. (2019). Hostile attribution bias and aggression in adults—A systematic review. *Aggression and Violent Behavior*, *46*, 66–81.

Wagner, N. J., Mills-Koonce, W. R., Propper, C. B., Willoughby, M. T., Rehder, P. D., Moore, G. A., & Cox, M. J. (2016). Associations between infant behaviors during the face-to-face still-face paradigm and oppositional defiant and callous-unemotional behaviors in early childhood. *Journal of Abnormal Child Psychology*, *44*(8), 1439–1453.

Walters, G. D. (2020). Crime and social cognition: A meta-analytic review of the developmental roots of adult criminal thinking. *Journal of Experimental Criminology*, *342*, 1–25.

Wilkowski, B. M., Robinson, M. D., Gordon, R. D., & Troop-Gordon, W. (2007). Tracking the evil eye: Trait anger and selective attention within ambiguously hostile scenes. *Journal of Research in Personality*, *41*, 650–666.

Zajenkowska, A., Jankowski, K. S., Lawrence, C., & Zajenkowski, M. (2013). Personality and individual differences in responses to aggression triggering events among prisoners and non-prisoners. *Personality and Individual Differences*, *55*(8), 947–951.

Zajenkowska, A., Mylonas, K., Lawrence, C., Konopka, K., & Rajchert, J. (2014). Cross-cultural sex differences in situational triggers of aggressive responses. *International Journal of Psychology*, *49*(5), 355–363.

Zajenkowska, A., Prusik, M., Jasielska, D., & Szulawski, M. (2020). Hostile attribution bias among offenders and non-offenders: Making social information processing more adequate. *Journal of Community & Applied Social Psychology*

Zajenkowska, A., Prusik, M., Jasielska, D., & Szulawski, M. (2021). Hostile attribution bias among offenders and non-offenders: Making social information processing more adequate. *Journal of Community & Applied Social Psychology*, *31*(2), 241–256.

Zajenkowska, A., Prusik, M., & Szulawski, M. (2018). What does the Ambiguous Intentions Hostility Questionnaire really measure? The

importance of context in evaluating hostility bias. *Journal of Personality Assessment*, *102*, 205–213.

Zajenkowska, A., & Rajchert, J. (2020). How sensitivity to provocation shapes encoding and interpretation of ambivalent scenes in an eye tracking study. *Journal of Cognitive Psychology*, *32*, 180–198.

Zajenkowska, A., Rogoza, R., Sasson, N. J., Harvey, P. D., Penn, D. L., & Pinkham, A. E. (2021). Situational context influences the degree of hostile attributions made by individuals with schizophrenia or autism spectrum disorder. *British Journal of Clinical Psychology*, *60*(2), 160–176.

How to cite this article: Zajenkowska, A. M., B odecka, M., Duda, E., & Lawrence, C. (2022). Reduced attention toward faces, intentionality and blame ascription in violent offenders and community-based adults: Evidence from an eye-tracking study. *Aggressive Behavior*, 48, 264–274. https://doi.org/10.1002/ab.22018

Article

Sex Differences in Inmates: Anger, Sensitivity to Provocation and Family History of Imprisonment

International Journal of Offender Therapy and Comparative Criminology
1–16

DOI: 10.1177/0306624X211049189
journals.sagepub.com/home/ijo
SAGE

Marta Bodecka-Zych[1], Anna Zajenkowska[1], and Mary Bower Russa[2]

Abstract

Little research has explored the role of aggression, anger, and family history of incarceration as they relate to female offenders. The current study aimed to address this gap in the literature by investigating these possible risk factors for incarceration among both men and women. The survey involved 123 (61 female and 62 male) prisoners convicted for violent crimes and a comparison group of 118 (60 female and 58 male) adults from the community. We found that women (convicted and non-convicted) were more sensitive to provocation than men, while community adults showed higher levels of trait anger than prisoners. Detainees were more likely than community adults to have a relative in prison. Although male and female inmates were equally likely to have a relative in prison, they differed in their relation to the imprisoned relative. Male and female prisoners showed increased risk for incarceration of same sex, first degree relatives (father and brothers for men, and mothers for women). These results may contribute to improved understanding of incarcerated populations. As such, this represents a critical first step in creating recovery programs that are more gender appropriate.

Keywords

imprisonment, trait anger, sensitivity to provocation, family risk, gender differences

While women's representation in prison settings continues to be relatively small (Eurostat Statistic Explained, 2019; Prison Service, 2019), there has been a vivid pattern of increasing representation of women in the prison system in the past few

[1]The Maria Grzegorzewska University, Warsaw, Poland
[2]Grand Valley State University, Allendale, MI, USA

Corresponding Author:
Marta Bodecka-Zych, The Maria Grzegorzewska University, Szczęśliwicka 40, Warsaw 02-353, Poland.
Email: mbsd1@aps.edu.pl

decades. In fact, the probability of imprisoning a female compared to a male has doubled from 1980 to 2008 (Cauffman, 2008), yet females are often not a significant focus in criminological research (Campaniello, 2019). Difficulties conducting prison research, combined with a low rate of violent female inmates has led to a literature on female prisoners that is, at times, incomplete, chaotic, and fragmented (Miszewski, 2016).

Questions about personal and environmental factors related to female crime are among the issues that remain less resolved (Hollin & Palmer, 2006). Traditional policies to reduce crime have often not distinguished between women and men (Campaniello, 2019; Suter et al., 2002) and prison programs and services designed for males are often extended to female inmates with little alteration (Koons et al., 1997). This is particularly problematic in light of the fact that both etiological contributors (e.g., Asscher et al., 2015) and the needs of males and females may be very different (e.g., Cauffman, 2008; Fields & Abrams, 2010; Hornsveld et al., 2018; Muller & Kempes, 2016). For instance, a recent study of male and female prisoners in the UK found that women report greater mental health needs in relation to men (e.g., personality disorders, mood disorders, PTSD; Tyler et al., 2019). Improved understanding of female inmates could improve attempts at intervention in this population.

In considering risk factors that may contribute to recidivism for male and female inmates, previous research among female prisoners convicted of violent crimes indicates that women have lower levels of hostility than men, but higher levels of anger (Hornsveld et al., 2018). Similar results, although from a sample involving relatively few violent female participants (12% of females were violent in contrast to 54% of males who were violent), were obtained by Suter et al. (2002). They identified inmate gender differences with regard to both experience and expression of anger. In general, their research using State Trait Anger Expression Inventory (STAXI, Spielberger, 1991), found that women received higher scores than men on state anger, trait anger, angry reaction, angry temperament, anger in, anger out and anger expression. Suter et al. (2002) notes that a similar pattern, with females scoring higher on anger or anger manifestation, has been found in clinical samples. Importantly, research to date suggests that women and men from the community do not differ in terms of trait anger (Archer, 2004; Averill, 1983; Deffenbacher et al., 1996).

Anger has been a considerable focus in the offender literature (e.g., Hornsveld et al., 2018; Lim et al., 2011; Roberton et al., 2015) and it is a risk factor for violent recidivism (Loza & Loza-Fanous, 1999). Studies to date suggest that the prison population is characterized by higher levels of anger than is found in the general population (Schönenberg & Jusyte, 2014; Suter et al., 2002), with level of anger in prison settings approaching levels found in clinical samples (e.g., Jones et al., 1999). Anger is a predictor of aggressive behavior (Smith & Waterman, 2006) and plays a key role in hostile or retaliatory violence (Browne & Howells, 1996; Holbrook, 1997). Anger is associated with increased aggressive behavior under conditions of provocation (Bettencourt et al., 2006), and provocation itself is among the most important causes of aggression (Anderson & Bushman, 2002).

Sensitivity to provocation may also influence the intensity of the aggressive responding (Lawrence & Hutchinson, 2013) Among prisoners, sensitivity to provocation is negatively associated with agreeableness (Zajenkowska et al., 2013), suggesting a tendency to lack trust, to be hostile and irritable, and to have a low regard for others (Costa et al., 1989). Such individuals may also exhibit destructive conflict resolution tactics, such as physical action and threats (Jensen-Campbell & Graziano, 2001). This pattern is consistent with a general tendency for individuals who are prone to provocation to misinterpret social situations as hostile and threatening (Bondü & Richter, 2016), although this tendency is more typical for males than for females (Zajenkowska & Rajchert, 2020).

A rich literature has examined the tendency for children whose parents are engaged in criminal activities to be more likely to be engaged in criminal activities themselves (e.g., Andrews & Bonta, 2010; Hawkins, 2000). Indeed, having a parent imprisoned before turning 10 years old makes it twice as likely that a male youth will commit a crime involving physical assault (Farrington, 1989). In a recent meta-analysis of 23 papers with nearly 3.5 million children, Besemer et al. (2017) confirmed the tendency for individuals with criminal behavior to be more likely to have parents with criminal behavior, with the pattern consistent across more than eight different countries. A range of theoretical explanations for this pattern have been proposed, including social learning theory, environmental influences, biological factors and others (see Besemer et al., 2017).

In fact, it seems most likely that criminal behavior and risk for incarceration is the result of interactions among a host of personal, environmental, and social factors. Anger and sensitivity to provocation may increase the tendency for aggressive responding in a parent (Lawrence, 2006) leading to difficulties with attachment with a child (Bowlby, 1973) and a tendency for overly punitive parenting approaches. In turn, the child may acquire genetic characteristics from the parent that place them at a similar risk for anger and aggression or antisocial behavior (e.g., Raine, 2002), while the child's experience of aggressive and overly punitive parenting approaches creates an opportunity for simultaneous social learning of aggression (e.g., Bandura, 2073; Farrington, 2011) and maladaptive coping and problem solving. Finally, some parents and their children (who later becomes an adults) are more likely to find themselves in environments and circumstances (e.g., poverty, unsafe neighborhoods) that encourage criminality in both the parents and offspring (Farrington, 2011).

Possible gender specific patterns in this intergenerational pattern still are not well understood. For example, early work focused primarily on risk for male youth (Farrington, 1989), with some indication that imprisonment of a father might have the greatest impact (Baker & Mednick, 1984). Farrington found that imprisonment of a father was predictive of later criminality of a son regardless of arrest circumstances for other relatives (Farrington et al., 2001). Similar research regarding females has lagged a bit (Farrington et al., 2001; Johnson, 1987), but early work has suggested that female inmates are more likely to have convicted mothers than convicted fathers (Farrington et al., 2001). In Besemer et al. (2017) recent meta-analysis, transmission appeared to be strongest from mothers to daughters, then mothers to sons and fathers to daughters, and lowest from fathers to sons.

In summary, prior research has shown that anger and associated vulnerability to provocation, as well as having an incarcerated relative, are risk factors for incarceration. However, details regarding the extent to which these variables play a role for male versus female inmates remains less clear. The aim of the current study is to clarify these uncertainties.

Current Study

The current study evaluated characteristics of female and male detainees in prisons using comparison to female and male individuals in a community sample. Because female inmates are more likely to be convicted for less violent crimes (Campaniello, 2019), we focused only on female and male prisoners convicted for violent offenses such as physical assault, attempted murder, or murder. This approach reduced the potential for confounds related to severity of the crime when making systematic comparisons between male and female inmates. Due to logistics of conducting research within a prison setting, all inmates who met eligibility criteria and were willing to participate were included in the research sample.

Violent crime perpetration may be associated with higher rates of psychopathology (e.g., Hornsveld et al., 2018; Suter et al., 2002), particularly among females, who present higher level of anger than men (Suter et al., 2002). Given this pattern, we predicted that violent female prisoners would present higher levels of trait anger as compared to male prisoners. We did not expect sex differences for non-prisoner (community) adults.

Although it has received very little attention in research with inmates, sensitivity to provocation may be a precursor and risk factor for anger and aggression (Bondü & Richter, 2016; Lawrence, 2006); thus, this variable was a secondary focus in the present study. While research on sensitivity to provocation has produced mixed results with regard to sex differences, there is some evidence suggesting gender differences in brain activation in situations of provocation. It also appears that aggressive men, when provoked, act more automatically than aggressive women (Repple et al., 2018). A meta-analysis on aggression as a function of provocation showed that unprovoked males are more aggressive than females, but when provoked, the differences decrease (Bettencourt & Miller, 1996). Our own studies indicate that women may show greater sensitivity to provocation than men, and it appears that this pattern might vary as a function of cultural features of the country in which the research is conducted (Lawrence, 2006; Zajenkowska et al., 2014). In general, female students have shown higher sensitivity to provocation (SP) than males (Zajenkowska et al., 2013, 2014) but similar research with prisoners found no sex-differences (Zajenkowska et al., 2013). In the current study we examined patterns of sensitivity to provocation in male and female inmates matched to a community sample. Based on past research using the same measure (Zajenkowska et al., 2013, 2014), we expected greater sex differences in SP among adults from the community.

Finally and most importantly, we wanted to compare history of a relative in prison among males and females in both prisoners and a non-prisoner sample to determine

Table 1. Demographics for Offender and Community Samples.

Characteristic	Prisoners ($n = 123$)	Non-prisoners ($n = 118$)
Age in years: *M* (*SD*)	37.32 (9.76)	35.61 (12.09)
Level of education: *N* (%)		
Completed primary school	48 (39)	11 (9.3)
Completed professional school	43 (35)	34 (28.8)
Completed high school	24 (19.5)	44 (37.3)
Beyond secondary education	4 (3.3)	29 (24.6)
Other	2 (1.6)	0

Note. Information regarding age was missing from four prisoners (3.3%). Information regarding education levels were missing from two prisoners (1.6%).

whether patterns mirrored those found in past research and whether gender differences in this potential risk factor emerged.

Method

Note on Data

The study was part of a larger research project, for which underlying data can be accessed at apsycholab.pl. All analyses and results reported here are novel. Further reports regarding sensitive to provocation, anger, and their relation to rejection sensitivity are reported in another work by Bodecka and colleagues (2021).

Participants

Participants consisted of 123 male and female inmates from two adult prisons in Poland. Investigators matched this sample to a group of 118 non-prisoners (i.e., "community sample") based on gender and age. Corrections officers, on the basis of judicial documents, identified persons who were convicted of violent crimes (murders, attempted murders, brawls, robberies) and invited them to participate in the study. Individuals who agreed to take part in the survey[1] were scheduled to meet with the researcher. There were no age differences between the prisoners and the community group $t(224.25) = 1.20$, $p = .23$), but groups significantly differed on education (see Table 1), as outside prison, it was difficult to recruit older individuals with lower educational backgrounds to the study. Low education is one of the risk factors for imprisonment (e.g., Andrews & Bonta, 2010; Hawkins, 2000); both community and prison samples appeared to be representative of the larger population from which they were selected.

All prison participants were Polish, and most of the prisoners lived in towns or cities (92% female, 76% male) before imprisonment. Some of them were in relationships (38% female, 26% male), while others were single (33% female, 54% male), and some had experienced broken relationships (divorce or death of the spouse—29% female,

Table 2. Conviction Statistics for Female and Male Inmates.

Prisoners	Female ($n = 61$)	Male ($n = 62$)
Age at first conviction[a]: *M* (*SD*)	30.48 (10.19)	22.55 (8.36)
Number of convictions[b]: *M* (*SD*)	1.66 (0.83)	2.71 (0.69)
Number of convictions: *N* (%)		
One conviction	35 (57.4)	2 (3.2)
Two convictions	12 (19.7)	17 (27.4)
Three or more convictions	14 (23)	43 (69.4)

Note. Information regarding age of first conviction was missing from one female.
[a]Groups differ significantly at $p < .001$.
[b]Groups differ significantly at $p < .001$.

20% male). For inmates, the average age was 37.32 years ($SD = 9.76$l, range 19–65 years). There were no significant differences with regard to age as a function of sex t (116.07) = 0.457, $p = .649$. Women had fewer prior convictions (t (115.97) = 7.64, $p < .001$) than men, and they were significantly older when they were first convicted (t (120) = 4.71, $p < .001$; see Table 2).

The community sample had 118 participants (60 female). The average age was 35.61 years ($SD = 12.08$, range 23–80 years); there were no sex differences with regard to age t (110.59) = −0.65, $p = .52$. Individuals from this group were recruited via social media or asked directly by investigators in their community to take part in a larger study. Most of the community sample, like prisoners, came from a town or city (68% female, 88% male), and there was also diversity in terms of close relationships (48% female and 41% male were single; 43% female and 57% male were in a relationships; 9% female and 2% male had a broken relationship).

Measures

Participants in the study completed a series of measures as part of a larger research project. Sensitivity to Provocation was measured using The Situational Triggers of Aggressive Responses (STAR) scale (Lawrence, 2006). The full questionnaire consists of 22 items (10 reflecting Sensitivity to Frustration—SF and 12 reflecting Sensitivity to Provocation—SP). For the purpose of this study we used only the SP subscale. Items included: a friend betrays me; I am the subject of a practical joke; someone insults me, someone behaves in an inconsiderate manner toward me, someone makes offensive remarks to me. Participants provided a rating of how aggressive each of the 12 situations typically made them feel using a 5-point, Likert scale. This instrument has high internal consistency ($\alpha = .82$ for SP) and has demonstrated validity in past research (Lawrence, 2006). It has also been used successfully in previous research in Poland (e.g., Zajenkowska & Konopka, 2015; Zajenkowska et al., 2013). Cronbach's alpha in the present research was $\alpha = .89$. Anger was assessed using The State-Trait Anger Expression Inventory—STAXI-2 (Spielberger & Reheiser, 2003),

which is a well-established measure of anger that has been used extensively in past research (Howells et al., 2005; Roberton et al., 2015; Velotti et al., 2017). The STAXI-2 contains 57 items comprising three scales (state anger, trait anger, and anger expression/control). Only the trait anger section was used in this study. This scale uses 10 items to assess how frequently angry feelings are experienced over time. Likert responses ranged from 1 (*almost never*) to 4 (*almost always*). The STAXI-2 has been validated in a forensic sample (Etzler et al., 2014) and adapted for use in Poland, where the reliability of each scale has been shown to be high (on average $\alpha = .80$; Bąk, 2016). Analyses confirmed similar reliability in the present study ($\alpha = .89$).

Information about family members was collected through questions: *"Has anyone in your family ever been convicted?"* and "*Who was that person to you*?"

Procedure

The study used a quasi-experimental research design. All participants completed structured and standardized self-report questionnaires. Individual measurements were conducted in designated rooms of the facility (in prison, it was usually a common room) by trained, senior psychology students from the research team. Participants who were interested in participating were provided an overview of the study and offered the opportunity to participate. If they chose to participate, they were informed that they could discontinue without penalty at any time. Participants were also informed that their responding would be confidential and that their responses would not be revealed to anyone. Upon completion of the questionnaires, participants were debriefed and thanked for their participation. The research study was approved by the university ethics committee and the prison directors.

Results

To determine whether prisoners and non-prisoners differed in terms of sensitivity to provocation and anger, a MANOVA was conducted. For this analysis, group (prisoners and non-prisoners) and gender were included as the between-subject variables, and sensitivity to provocation (STAR) and anger (STAXI-2)[2] were the within-subject variables. The analysis revealed a main effect of gender in case of STAR $F(1, 237) = 10.54$, $p = .001$, $\eta^2_p = .043$, such that women ($M = 3.77$, $SD = 0.08$) were more sensitive to provocation than men ($M = 3.43$, $SD = 0.08$); the main effect of prisoner versus non-prisoner group was not significant. In the case of Anger (STAXI-2), analyses were significant $F(1, 237) = 3.92$, $p = .49$, $\eta^2_p = .016$, with non-prisoners ($M = 2.48$, $SD = 0.06$) showing higher levels of trait anger than prisoners ($M = 2.30$, $SD = 0.06$). The interaction effect between group and gender was non-significant in the case of both STAXI-2 ($p = .76$) and STAR ($p = .61$).

A chi-square test of independence was performed to examine the relation between group membership (prisoners and non-prisoners) and likelihood of having an imprisoned relative. The relation between these variables was significant, χ^2 $(1, 239) = 47.21$, $p < .001$. Prisoners were more likely than non-prisoners to have a relative in prison:

Table 3. Relatives in Prison of Female and Male Prisoners.

Prisoners	Female ($n = 61$)	Male ($n = 62$)
Relative in prison*: *N* (%)		
Father	9 (14.8)	14 (23.3)
Mother	8 (13.1)	0
Father and mother	1 (1.6)	1 (1.7)
Brother	5 (8.2)	13 (21.7)
Other (e.g., uncle)	2 (3.3)	3 (5)
Many (more than 2)	2 (3.3)	1 (1.7)
No one	34 (55.7)	28 (46.7)

Note. Information regarding relatives in prison was missing from two male.
*The categories are separable.

48.8% of inmates had a family member (usually a father) in prison, compared to 8.5% of the non-prisoner population. When gender of inmate was considered, women had a relative in prison as often as men, χ^2 $(1, 121) = 0.996$, $p = .318$, however female prisoners and male prisoners differed in their relation to the imprisoned relative, χ^2 $(6, 121) = 13.75$, $p = .033$. As shown on Table 3, 15% of women in prison had imprisoned fathers and 13% had imprisoned mothers; while 23% of men had imprisoned fathers, and no men had imprisoned mothers.

Discussion

The present study focused on examining psychological and environmental characteristics of female and male prisoners that may be associated with risk for recidivism, and comparing these characteristics to those of females and males in a community sample. Both community and inmate females showed greater sensitivity to provocation than males, but there were no sex differences in trait anger. The community sample showed higher levels of anger than prisoners. With regard to family history of incarceration, detainees were more likely than adults from the general population to have a relative in prison. While female and male inmates were equally likely to have a relative in prison, female inmates' and male inmates' relationship to the imprisoned relative differed: females were more likely to have an imprisoned mother or father, while males were more likely to have an imprisoned father or brother.

Our findings highlight the fact that sensitivity to provocation and trait anger are related but not identical constructs (Bondü & Richter, 2016). Individuals who are sensitive to provocation react strongly to external triggers, and those triggers may result in feelings of anger and reactive aggression (Lawrence, 2006). In turn, individuals characterized by a heightened anger trait experience a state of anger more frequently and more strongly, but this is not necessarily situation related, and it may be due to temperament (Bąk, 2016).

Our findings regarding gender differences with regard to these two characteristics are largely consistent with patterns reported in previous research. For example, the

finding that females are more sensitive to provocation than males is consistent with previous research where female students (e.g., in Poland, Greece) were more likely than male students to report feeling aggressive as a result of direct provocation (Zajenkowska et al., 2014). Similar patterns have also appeared in research on aggressive feelings and beliefs in violent inmates in the UK (Archer & Haigh, 1997). On the other hand, the lack of sex differences in terms of anger in inmates convicted of violent crime are consistent with previous studies finding no gender differences in the frequency, intensity and duration of the reported anger in general population (e.g., Archer, 2004).

Previous research on sensitivity to provocation among prisoners did not find the sex difference that were apparent in our present study (Zajenkowska et al., 2013). Differences may reflect variations in the prisoner samples, since this current study (unlike the previous one) focused only on the most violent female and male inmates. Future research may wish to explore the possibility that gender and resulting social roles may be more important predictors of differences in trait anger and sensitivity to provocation than having been convicted of a violent crime.

The pattern of sex differences with anger may require further exploration. It is possible that, particularly for women, "anger" is also expressed or experienced in other closely-related, yet somewhat distinct, emotional forms (e.g., irritation, frustration, annoyance). In fact, narrative studies with teenage men and women show sex differences in anger perception. Unlike male teenagers, teenage women often see anger as a "burden" (Budziszewska & Hansen, 2019). Our data raise the possibility that the variability in the experience of angry feelings in reaction to provocation within gender could be greater than between genders, since individuals may subjectively experience anger very differently (Shields, 2013).

Surprisingly, prisoners in the present study were less angry than adults from the community sample (the result is on the verge of a statistical trend). It is possible that, due to the prison context in which the study took place, inmates experienced a process of habituation to a variety of provocation events, which led to a lower level of experienced anger. It is also possible that the prison setting induces a sense of lack of control (i.e., learned helplessness), resulting in a lower level of the motivation or drive that is often the source of anger (Harmon-Jones et al., 2013). The possibility that this leads to reduced anger and frustration in inmates (e.g., Carver & Harmon-Jones, 2009) needs further attention.

We cannot rule out the possibility that the low ratings of anger among inmates may have been influenced by measurement factors. The State-Trait Anger Expression Inventory - STAXI-2 (Spielberger & Reheiser, 2003), although widely used in forensic research is considered susceptible to impression management. McEwan et al. (2009) have demonstrated that socially desirable response bias may be associated with lower reported anger in forensic samples. In spite of this, self-report is typically used to assess anger in offender populations (e.g., Hornsveld et al., 2018; Howells et al., 2005; Roberton et al., 2015), and assurances that responses would be confidential may have served to attenuate tendencies for socially desirable responding in the present study. However, future studies on anger in prisons may benefit from the inclusion of physiological measures (e.g., galvanic skin response or pupil dilation) which may be less amenable to socially desirable responding.

As expected, prisoners were more likely than non-prisoners to have a relative in prison. Our data are consistent with research that crime is often nested within families, so that incarceration of one family member increases the likelihood that another family member will also experience incarceration (e.g., Farrington et al., 2001; Kiliszek, 2013). The increased family history of incarceration among inmates may reflect a tendency for social learning of aggressive behavior (Bandura, 1973; Lefkowitz et al., 1977) from parent figures in the home. It is also possible that transfer of imprisonment is associated with the development of a tendency to hostile attributions that increase the risk of aggressive behavior (Crick & Dodge, 1994) and that are acquired during interaction with the caregiver and during socialization (Dodge, 2006). Finally, it is clear that the probability of a child being imprisoned can be increased by the absence of the father in the family home (Harper & McLanahan, 2004), which could be related to loss of security, lower self-esteem or poverty (e.g., O'Neill, 2002).

The tendency for male and female prisoners to show increased risk for incarceration of same sex, first degree relatives (father and brothers for men and mothers for women) may reflect an increased tendency for children to model the behavior of the same-sex parent. Consistent with this notion, a good relation with the same-sex parent can be a protective factor against criminal activity (Laible & Carlo, 2004). Indeed, fathers may play a critical role in shaping the trajectories for children of both genders: Hoeve et al. (2009), provide evidence that poor paternal support is more strongly related to child criminality than poor maternal support. However, fathers appear to have more influence on their sons' delinquency than mothers. Similarly, with regard to interventions, parental training sessions in which the father participates are more effective than those in which the mother is alone (Lundahl et al., 2008).

Limitations and Future Research

While attempts were made to match our group of prisoners to a community sample for comparison, there were educational differences between the group of inmates and the community sample. The prisoner sample consisted of a number of individuals with primary or vocational education, which is consistent with the tendency for low educational level to be one of the most significant risk factors for imprisonment (e.g., Wasserman, 2003). Unfortunately, it proved difficult to recruit adults with limited education for the community sample. As a result, differences between prisoner and non-prisoner groups (e.g., family history of incarceration) may have been influenced by educational level. Additionally, we did not receive information from the prison authorities regarding the rates at which inmates declined to take part in the study. It is possible that individuals with specific psychological characteristics (e.g., higher agreeableness) were more likely to participate, so attention to such sampling characteristics would be important in future research.

A second limitation was that our reliance on self-report of anger may have led to underreporting of anger this inmate population. Although this possibility cannot be ruled out completely, research using the same self-report instrument found higher rates

of anger for incarcerated offenders in Italy and Australia than those living in the community on parole (Velotti et al., 2017). While cultural differences make direct comparisons difficult, this finding attenuates some of the concern that inmates may be simply unwilling to report anger. Additionally, violent versus nonviolent inmates have shown expected differences in anger scores on this instrument in past research, a finding that further supports the validity of this instrument for use in the prison population (Lim et al., 2011). Future work with prisoners may consider the possibility of using physiological responding or qualitative approaches to better understand anger in this population.

Conclusion and Policy Implications

Despite its limitations, our study highlights several important findings. First, it appears that differences in trait anger and sensitivity to provocation between prisoners and the community are not as large as commonly believed. Often differences between inmates and non-inmates are assumed a priori, but there are not many studies that compare forensic groups with the general population. Additional research comparing mechanisms of action of individual traits for inmates and non-inmates, as well as their interactive effects with the situational context, would be useful. While we found patterns by which women (both inmates and non-inmates) showed stronger sensitivity to provocation than men, the implications of this pattern in offender populations remain unclear. The possible impacts of this sensitivity for social dynamics of female prisoners both while incarcerated and upon release, warrants further consideration. Additionally, to the extent that decreased anger in inmate populations (compared to community samples) might reflect a pattern of learned helplessness, this raises concerns about whether inmates who are incarcerated maintain the degree of self-efficacy that may be required for successful re-engagement in society and re-integration to the community upon release. Finally, further exploration of the impacts of parent imprisonment on socialization and development of children within the home are warranted. For example, our data suggest that prevention programs might benefit from increased attention to incarcerated first degree relatives of inmates, who may find themselves on the fringes and at risk of entry to the criminal justice system.

In conclusion, continued work to identify unique characteristics and needs of male and female inmates can only improve our ability to meet the needs of these offender populations via primary and secondary prevention, as well as during incarceration and upon release. As our understanding of the complex and interactive processes that lead to criminal behavior and incarceration improve, intervention programing can better target the ways in which the family histories, social experiences, and personal characteristics of male and female prisoners shape their developmental trajectories (e.g., Dixon et al., 2004; Roberton et al., 2015).

Data Availability

Data used in this study will be stored at: http://www.apsycholab.pl

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Science Centre under Grant [UMO-2017/26/D/HS6/00258].

ORCID iDs

Marta Bodecka-Zych https://orcid.org/0000-0001-7766-1446

Anna Zajenkowska https://orcid.org/0000-0002-9413-1428

Notes

1. Recruitment of inmates was conducted by prison staff; information regarding number of refusals to participate were not provided to us.
2. Missing data were replaced using the average of the observations.

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

Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct* (5th ed.). Routledge.

Archer, J. (2004). Sex differences in aggression in real-world settings: A meta-analytic review. *Review of General Psychology*, *8*(4), 291–322. https://doi.org/10.1037/1089-2680.8.4.291

Archer, J., & Haigh, A. (1997). Beliefs about aggression among male and female prisoners. *Aggressive Behavior*, *23*(6), 405–415. https://doi.org/10.1002/(sici)1098-2337(1997)23:6<405::aid-ab1>3.0.co;2-f

Asscher, J. J., Van der Put, C. E., & Stams, G. J. J. (2015). Gender differences in the impact of abuse and neglect victimization on adolescent offending behavior. *Journal of Family Violence*, *30*(2), 215–225. https://doi.org/10.1007/s10896-014-9668-4

Averill, J. R. (1983). Studies on anger and aggression: Implications for theories of emotion. *American Psychologist*, *38*, 1145–1160. https://doi.org/10.1037/0003-066X.38.11.1145

Bąk, W. (2016). Pomiar stanu, cechy, ekspresji i kontroli złości. Polska adaptacja kwestionariusza STAXI-2 [The measurement of state, expression and control of anger. Polish adaptation of STAXI-2 questionnaire]. *Polskie Forum Psychologiczne*, *21*(1), 93–122. https://doi.org/10.14656/PFP20160107

Baker, R. L., & Mednick, B. R. (1984). *Influences on human development: A longitudinal analysis*. Mass Kluwer Academic Publishers.

Bandura, A. (1973). *Aggression: A social learning analysis*. Holt.

Besemer, S., Ahmad, S. I., Hinshaw, S. P., & Farrington, D. P. (2017). A systematic review and meta-analysis of the intergenerational transmission of criminal behavior. *Aggression and Violent Behavior*, *37*, 161–178. https://doi.org/10.1016/j.avb.2017.10.004

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

Bettencourt, B. A., Talley, A., Benjamin, A. J., & Valentine, J. (2006). Personality and aggressive behavior under provoking and neutral conditions: A meta-analytic review. *Psychological Bulletin*, *132*(5), 751–777. https://doi.org/10.1037/0033-2909.132.5.751

Bodecka, M., Grala, K., Szymczak, K., Zajenkowska, A., Rajchert, J., Kostrzewska, D., & Lawrence, C. (2021). Relationship between trait anger and rejection sensitivity: The moderating role of the group (inmates and non-inmates). *The Journal of Forensic Psychiatry & Psychology*. Advance online publication. https://doi.org/10.1080/14789949.2021.1909104

Bondü, R., & Richter, P. (2016). Interrelations of justice, rejection, provocation, and moral disgust sensitivity and their links with the hostile attribution bias, trait anger, and aggression. *Frontiers in Psychology*, *7*, 795. https://doi.org/10.3389/fpsyg.2016.00795

Bowlby, J. (1973). *Attachment and loss: Volume II: Separation, anxiety and anger*. The Hogarth Press and the Institute of Psycho-Analysis.

Browne, K., & Howells, K. (1996). Violent offenders. In C. R. Hollin (Ed.), *Working with offenders: Psychological practice in offender rehabilitation* (pp. 188–210). Wiley.

Budziszewska, M., & Hansen, K. (2019). "Anger detracts from beauty": Gender differences in adolescents' narratives about anger. *Journal of Adolescent Research*, *35*(5), 635–664. https://doi.org/10.1177/0743558419845870

Campaniello, N. (2019). *Women in crime*. IZA World of Labor, Institute of Labor Economics (IZA), Bonn, Iss.105v2. https://doi.org/10.15185/izawol.105.v2

Carver, C. S., & Harmon-Jones, E. (2009). Anger is an approach-related affect: Evidence and implications. *Psychological Bulletin*, *135*, 183–204. https://doi.org/10.1037/a0013965

Cauffman, E. (2008). Understanding the female offender. *The Future of Children*, *18*(2), 119–142. https://doi.org/10.1353/foc.0.0015

Costa, P. T., McCrae, R. R., & Dembroski, T. M. (1989). Agreeableness versus antagonism: Explication of a potential risk factor for CHD. In A. W. Siegman & T. M. Dembroski (Eds.), *In search of coronary prone behavior* (pp. 41–63). Erlbaum.

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin*, *115*(1), 74.

Deffenbacher, J. L., Oetting, E. R., Thwaites, G. A., Lynch, R. S., Baker, D. A., Stark, R. S., & Eiswerth-Cox, L. (1996). State–trait anger theory and the utility of the trait anger scale. *Journal of Counseling Psychology*, *43*(2), 131–214. https://doi.org/10.1037/0022-0167.43.2.131

Dixon, A., Howie, P., & Starling, J. (2004). Psychopathology in female juvenile offenders. *Journal of Child Psychology and Psychiatry*, *45*, 1150–1158. https://doi.org/10.1111/j.1469-7610.2004.00307.x

Dodge, K. A. (2006). Translational science in action: Hostile attributional style and the development of aggressive behavior problems. *Development and Psychopathology*, *18*, 791–814. https://doi.org/10.1017/S0954579406060391

Etzler, S. L., Rohrmann, S., & Brandt, H. (2014). Validation of the STAXI-2: A study with prison inmates. *Psychological Test and Assessment Modeling*, *56*(2), 178.

Eurostat Statistic Explained. (2019, October). *Prison statistics*. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Prison_statistics

Farrington, D. P. (1989). Early predictors of adolescent aggression and adult violence. *Violence and Victims*, *4*(2), 79–100. https://doi.org/10.1891/0886-6708.4.2.79

Farrington, D. P. (2011). Families and crime. In J. Q. Wilson & J. Petersilia (Eds.), *Crime and Public Policy* (pp. 130–157). Oxford University Press.

Farrington, D. P., Jolliffe, D., Loeber, R., Stouthamer-Loeber, M., & Kalb, L. M. (2001). The concentration of offenders in families, and family criminality in the prediction of boys' delinquency. *Journal of Adolescence*, *24*(5), 579–596. https://doi.org/10.1006/jado.2001.0424

Fields, D., & Abrams, L. S. (2010). Gender differences in the perceived needs and barriers of youth offenders preparing for community reentry. *Child & Youth Care Forum*, *39*(4), 253–269. https://doi.org/10.1007/s10566-010-9102-x

Harmon-Jones, E., Gable, P. A., & Price, T. F. (2013). Does negative affect always narrow and positive affect always broaden the mind? Considering the influence of motivational intensity on cognitive scope. *Current Directions in Psychological Science*, *22*(4), 301–307. https://doi.org/10.1177/0963721413481353

Harper, C. C., & McLanahan, S. S. (2004). Father absence and youth incarceration. *Journal of Research on Adolescence*, *14*(3), 369–397. https://doi.org/10.1111/j.1532-7795.2004.00079.x

Hawkins, J. D. (2000). *Predictors of youth violence*. US Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention.

Hoeve, M., Dubas, J. S., Eichelsheim, V. I., Van der Laan, P. H., Smeenk, W., & Gerris, J. R. (2009). The relationship between parenting and delinquency: A meta-analysis. *Journal of Abnormal Child Psychology*, *37*(6), 749–775. https://doi.org/10.1007/s10802-009-9310-8

Holbrook, M. I. (1997). Anger management training in prison inmates. *Psychological Reports*, *81*(2), 623–626. https://doi.org/10.2466/PR0.81.6.623-626

Hollin, C. R., & Palmer, E. J. (2006). Criminogenic need and women offenders: A critique of the literature. *Legal and Criminological Psychology*, *11*(2), 179–195. https://doi.org/10.1348/135532505X57991

Hornsveld, R. H. J., Zwets, A. J., Leenaars, E., Kraaimaat, F. W., Bout, R., Lagro-Janssen, T., & Kanters, T. (2018). Violent female offenders compared with violent male offenders on psychological determinants of aggressive behavior. *International Journal of Offender Therapy and Comparative Criminology*, *62*, 450–467. https://doi.org/10.1177/0306624X16648109

Howells, K., Day, A., Williamson, P., Bubner, S., Jauncey, S., Parker, A., et al. (2005). Brief anger management programs with offenders: Outcomes and predictors of change. *Journal of Forensic Psychiatry and Psychology*, *16*, 296–311. https://doi.org/10.1080/14789940500096099

Jensen-Campbell, L. A., & Graziano, W. G. (2001). Agreeableness as a moderator of interpersonal conflict. *Journal of Personality*, *69*(2), 323–361. https://doi.org/10.1111/1467-6494.00148

Johnson, R. (1987). *Hard time: Understanding and reforming the prison*. Brooks/Cole.

Jones, J. P., Thomas-Peter, B. A., & Trout, A. (1999). Normative data for the Novaco Anger Scale from a non-clinical sample and implications for clinical use. *British Journal of Clinical Psychology*, *38*, 417–424. https://doi.org/10.1348/014466599163024

Kiliszek, E. (2013). Czynniki ryzyka sprzyjające niedostosowaniu społecznemu i przestępczości nieletnich. *Profilaktyka społeczna i resocjalizacja*, *21*, 165–222.

Koons, B. A., Burrow, J. D., Morash, M., & Bynum, T. (1997). Expert and offender perceptions of program elements linked to successful outcomes for incarcerated women. *Crime & Delinquency*, *43*(4), 5212–532. https://doi.org/10.1177/0011128797043004007

Laible, D. J., & Carlo, G. (2004). The differential relations of maternal and paternal support and control to adolescent social competence, self-worth, and sympathy. *Journal of Adolescent Research*, *19*(6), 759–782. https://doi.org/10.1177/0743558403260094

Lawrence, C. (2006). Measuring individual responses to aggression-triggering events: Development of the situational triggers of aggressive responses (STAR) scale. *Aggressive Behavior*, *32*(3), 241–252. https://doi.org/10.1002/ab.20122

Lawrence, C., & Hutchinson, L. (2013). The influence of individual differences in sensitivity to provocations on provoked aggression. *Aggressive Behavior*, *39*(3), 212–221. https://doi.org/10.1002/ab.21473

Lefkowitz, M. M., Eron, L. D., Walder, L. O., & Huesmann, L. R. (1977). *Growing up to be violent. A longitudinal study of the development of aggression*. Pergamon Press.

Lim, L., Day, A., & Casey, S. (2011). Social cognitive processing in violent male offenders. *Psychiatry Psychology and Law*, *18*, 177–189. https://doi.org/10.1080/13218711003739490

Loza, W., & Loza-Fanous, A. (1999). Anger and prediction of violent and nonviolent offenders recidivism. *Journal of Interpersonal Violence*, *14*, 1014–1029. https://doi.org/10.1177/088626099014010002

Lundahl, B. W., Tollefson, D., Risser, H., & Lovejoy, M. C. (2008). A meta-analysis of father involvement in parent training. *Research on Social Work Practice*, *18*(2), 97–106. https://doi.org/10.1177/1049731507309828

McEwan, T. E., Davis, M. R., MacKenzie, R., & Mullen, P. E. (2009). The effects of social desirability response bias on STAXI-2 profiles in a clinical forensic sample. *British Journal of Clinical Psychology*, *48*(4), 431–436. https://doi.org/10.1348/014466509X454886

Miszewski, K. (2016). Kobiety odbywające karę długoterminowego pozbawienia wolności. *Kwartalnik poświęcony zagadnieniom kryminologicznym i penitencjarnym*, *91*, 69–90.

Muller, E., & Kempes, M. (2016). Gender differences in a Dutch forensic sample of severe violent offenders. *International Journal of Forensic Mental Health*, *15*(2), 164–173. https://doi.org/10.1080/14999013.2016.1166464

O'Neill, R. (2002). *The fatherless family*. CIVITAS–The Institute for the Study of Civil Society.

Prison Service. (2019). *Krajowy rejestr karny* [National criminal record]. Prison Service.

Raine, A. (2002). Biosocial studies of antisocial and violent behavior in children and adults: A review. *Journal of Abnormal Child Psychology*, *30*(4), 311–326. https://doi.org/10.1023/A:1015754122318

Repple, J., Habel, U., Wagels, L., Pawliczek, C. M., Schneider, F., & Kohn, N. (2018). Sex differences in the neural correlates of aggression. *Brain Structure and Function*, *223*(9), 4115–4124. https://doi.org/10.1007/s00429-018-1739-5

Roberton, T., Daffern, M., & Bucks, R. S. (2015). Beyond anger control: Difficulty attending to emotions also predicts aggression in offenders. *Psychology of Violence*, *5*, 74–83. https://doi.org/10.1037/a0037214

Schönenberg, M., & Jusyte, A. (2014). Investigation of the hostile attribution bias toward ambiguous facial cues in antisocial violent offenders. *European Archives of Psychiatry and Clinical Neuroscience*, *264*(1), 61–69. https://doi.org/10.1007/s00406-013-0440-1

Shields, S. A. (2013). Gender and emotion: What we think we know, what we need to know, and why it matters. *Psychology of Women Quarterly*, *37*, 423–435. https://doi.org/10.1177/0361684313502312

Smith, P., & Waterman, M. (2006). Self-reported aggression and impulsivity in forensic and non-forensice populations: The role of gender and experience. *Journal of Family Violence*, *21*, 425–437. https://doi.org/10.1007/s10896-006-9039-x

Spielberger, C. D. (1991). *State-Trait anger expression inventory-revised research edition: Professional manual*. Psychological Assessment Resources.

Spielberger, C. D., & Reheiser, E. C. (2003). Measuring anxiety, anger, depression, and curiosity as emotional states and personality traits with the STAI, STAXI, and STPI. In M. J. Hilsenroth & D. L. Segal (Eds.), *Comprehensive handbook of psychological assessment* (Vol. 2, pp. 70–86). John Wiley & Sons Inc.

Suter, J. M., Byrne, M. K., Byrne, S., Howells, K., & Day, A. (2002). Anger in prisoners: Women 0RW1S34RfeSDcfkexd09rT2are1RW1S34RfeSDcfkexd09rT2 different from men. *Personality and Individual Differences*, *32*(6), 1087–1100. https://doi.org/10.1016/S0191-8869(01)00105-2

Tyler, N., Miles, H. L., Karadag, B., & Rogers, G. (2019). An updated picture of the mental health needs of male and female prisoners in the UK: Prevalence, comorbidity, and gender differences. *Social Psychiatry and Psychiatric Epidemiology*, *54*(9), 1143–1152. https://doi.org/10.1007/s00127-019-01690-1

Velotti, P., Garofalo, C., Callea, A., Bucks, R. S., Roberton, T., & Daffern, M. (2017). Exploring anger among offenders: The role of emotion dysregulation and alexithymia. *Psychiatry, Psychology and, Law*, *24*, 128–138. https://doi.org/10.1080/13218719.2016.1164639

Wasserman, G. A. (2003). *Risk and protective factors of child delinquency*. US Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention.

Zajenkowska, A., Jankowski, K. S., Lawrence, C., & Zajenkowski, M. (2013). Personality and individual differences in responses to aggression triggering events among prisoners and non-prisoners. *Personality and Individual Differences*, *55*(8), 947–951. https://doi.org/10.1016/j.paid.2013.07.467

Zajenkowska, A., & Konopka, K. (2015). Communion-orientation as an antidote for aggressive behaviour among high provocation sensitive individuals. *Personality and Individual Differences*, *76*, 62–67. https://doi.org/10.1016/j.paid.2014.11.048

Zajenkowska, A., Mylonas, K., Lawrence, C., Konopka, K., & Rajchert, J. (2014). Cross-cultural sex differences in situational triggers of aggressive responses. *International Journal of Psychology*, *49*(5), 355–363. https://doi.org/10.1002/ijop.12052

Zajenkowska, A., & Rajchert, J. (2020). How sensitivity to provocation shapes encoding and interpretation of ambivalent scenes in an eye tracking study. *Journal of Cognitive Psychology*, *32*(2), 180–198. https://doi.org/10.1080/20445911.2020.1717498

Motivation and Emotion
https://doi.org/10.1007/s11031-022-09972-z

ORIGINAL PAPER

Gender differences in sensitivity to provocation and hostile attribution bias toward ambiguous facial cues in violent offenders and community-based adults

Anna Zajenkowska[1] · Marta Bodecka-Zych[1] · Nina Gehrer[2] · Krzysztof Krejtz[3] · Claire Lawrence[4] · Michael Schoenenberg[2] · Aiste Jusyte[2]

Accepted: 10 July 2022

Abstract
Aggressive offenders commonly show hostile attribution bias in the perception of facial affect. Individuals' sensitivity to provocation has been also linked to hostile attribution. However, most studies have been limited to male offenders. The current study investigated whether sensitivity to provocation (SP) predicted bias towards interpretation of ambiguous facial cues as angry (hostility bias) in violent inmates compared to community-dwelling non-inmates. The sample (N = 272) consisted of 105 (53 women) violent inmates and 167 (85 women) adults living in the community. Hostility bias towards targets' faces was differently related to sensitivity to provocation across genders and groups depending on the target's sex. Generally, the higher inmates' sensitivity to provocation in men, the higher the identification of anger on female target faces, but the lower on male faces (anger/fear morphs). Conversely, the higher inmates' sensitivity to provocation in women, the *lower* the identification of anger on female target faces (anger/fear morphs). Additionally, we observed that in non-inmates, men's sensitivity to provocation significantly predicted anger identification on male faces. In the case of anger/happy face morphs, the more sensitive to provocation women in our samples were, the less they perceived anger in ambiguous faces. Conversely, men who were sensitive to provocation tended to perceive anger more often. With the current project, we show the importance of studying gender differences, which are often neglected, in the study of hostile interpretations of ambiguous stimuli amongst inmates and community samples. We anticipate that the results may help to design distinct and adequate resocialization and psychotherapeutic programs for both women and men with a tendency to violence.

Keywords Hostile attribution bias · Sensitivity to provocation · Anger · Inmates · Gender differences · Morphed faces

Introduction

"Hostile people tend to view the world as a hostile place" (Bushman, 2016, p.1). In particular, aggressive individuals are prone to interpret a peer's intention as hostile and also to interpret ambiguous nonverbal cues, like facial expressions, as hostile (Schönenberg & Jusyte, 2014). Hostile attributions in relation to the interpretation of a harmful social encounter is the belief that the actor did something with hostile purpose, intentionally, and is to be blamed for that, even though the social cues may be ambiguous or even neutral (Combs et al., 2007; Dodge, 2006; Tuente et al., 2019). Additionally, people may also present a hostile response bias in relation to more basic reactions, like emotion recognition. In this case, individuals will tend to give anger interpretations to ambiguous facial expressions (e.g., morphed faces consisting of 50% anger and 50% of fear, Schönenberg & Jusyte 2014). Facial expression carries a broad range of socially relevant information reflecting the internal states of the individual. Hence, processing facial affect is crucial for socialization and normal social interaction (Corden et al., 2006; Fridlund, 1991). A failure to be appropriately guided by such social cues is assumed to contribute to the development and maintenance of aggression and other maladaptive

✉ Anna Zajenkowska
zajenkowska@gmail.com; azajenkowska@aps.edu.pl

[1] Maria Grzegorzewska University, Szczęśliwicka 40, 00-712 Warsaw, Poland

[2] University of Tübingen, Tübingen, Germany

[3] SWPS University of Social Sciences and Humanities, Warsaw, Poland

[4] Lawrence PsychAdvisory, Nottingham, UK

Published online: 12 August 2022

antisocial behaviors (Blair, 2003; Montagne et al., 2005). For example, a study using morphed face stimuli in male violent offenders showed that aggressive behavior was associated with a strong preference to interpret ambiguous facial expressions as hostile (a hostile response bias was defined as the average number of anger interpretations of ambiguous facial expressions), but there was no evidence for a generally biased interpretation of distress cues (such as fear) under ambiguous conditions (Schönenberg & Jusyte, 2014). Misreading social interpersonal cues can lead to erroneously assuming the other person means to inflict harm. Such hostile interpretations can lead to highly-charged aggressive encounters, particularly for individuals with emotion regulation difficulties (Dodge et al., 1990). Further, a meta analytic review confirmed that, to some extent, the combination of anger and hostility bias when processing facial expressions is causally linked to aggressive behavior, especially in susceptible populations (Mellentin et al., 2015). Nonetheless, there is a current lack of research regarding facial affect bias, especially in aggression-prone and provocation-prone populations (Deffenbacher, 1992, Mellentin et al., 2015). Only few studies have dealt with the interpretation of ambiguous facial cues in antisocial populations (e.g., adult antisocial male offenders, Schönenberg & Jusyte, 2014) and none of the studies dealt with gender differences in violent offenders' samples in facial affect bias (Mellentin et al., 2015). However, such gender differences should not be ignored in affective research (Kret & De Gelder, 2012), especially in populations, such as violent inmates, where hostile perceptions may lead to aggression.

Importantly, there has been a significant increase of female inmates, which is at a much faster rate in comparison to males (Austin et al., 1993; Snell & Morton, 1994). Therefore there is a great need to analyze women's commission of violent crimes (Heidensohn, 1991; Verona & Carbonell 2000). At the same time, there is evidence for a link between women's assaultive or violent behaviors and the tendency to overcontrolled hostility (Ogle et al., 1995; Verona & Carbonell, 2000). Overcontrolled hostility occurs amongst individuals who, for whatever reason, typically suppress their angry or aggressive impulses, and then react to a provocation in a more extreme aggressive manner following a period of extended impulse control. One hypothesis to explain why overcontrolled aggression is seen in women is that the socialization of women leads to inhibition of aggressive impulses, which might result in accumulation of instigations and provocations and eventually in outburst of aggression when a threshold is surpassed (Ogle et al., 1995; Verona & Carbonell, 2000). Consequently, women may tend to pick up emotional cues from others that are inhibiting aggressive responses and rather facilitating a better flow of communication; whereas men would more easily perceive threatening cues, especially from other men (see for review Kret & De Gelder 2012). Men are superior to women at identifying anger (Mandal & Palchoudhury, 1985; Rotter & Rotter, 1988) and male aggressive offenders present a bias toward interpreting positive emotions as anger (Kret, 2011). Women on contrary outperform men in recognizing facial expressions of fear and sadness (Mandal & Palchoudhury, 1985). There are also differences in sensitivity to provocation between men and women.

Sensitivity to provocations (SP: the dispositional propensity for a person to perceive provocation more readily, and experience anger and aggressive feelings when provoked) is positively related to trait anger (the dispositional propensity for a person to experience angry emotions) and hostile attributions (Bondu & Richter, 2016; Schultz et al., 2010). Individuals sensitive to provocation typically report higher levels of anger, which is linked to both perceived intentionality of an act and perceived blame allocated to the harm-doer (Zajenkowska et al., 2021). Higher levels of sensitivity to provocation biases individuals to perceive others' behavior as inciting (Lawrence, 2006), and predicts higher aggression (Ayduk et al., 2008; Bondü & Krahé, 2015; Downey et al., 1998; Lawrence 2006). Interestingly while women from a general population tend to feel more aggressive in response to provocations than men (Zajenkowska et al., 2013), the more sensitive they are to provocation, the less they perceive intentionality in the ambiguous behavior of others (Zajenkowska & Rajchert, 2020). It is possibly that to avoid acting out their aggressive feelings and therefore being negatively evaluated by society, women tend to avoid conflictual situations (Ogle et al., 1995).

Lawrence & Hodgkins (2009) suggested that the influence of individual differences in sensitivity to provocation should be investigated rather than trait aggression alone when it comes to the interpretation of aggressive behavior. They demonstrated that individuals high on sensitivity to provocation interpret the behavior of a provoking individual to be more aggressive, and called for this effect to be examined further (Lawrence & Hodgkins, 2009). One way to find out why high SP individuals differ in the perception of harmful acts from low SP individuals is to examine individuals' encoding patterns of social cues (Zajenkowska & Rajchert, 2020). Individuals low on SP gaze longer at non-hostile cues than hostile cues. However, individuals high on SP do not focus their gaze significantly longer on either hostile or non-hostile cues in scenes depicting ambiguous social encounters, unless they are also high on trait anger. Those high in SP and trait anger gazed longer at the non-hostile cues compared to hostile cues (Zajenkowska & Rajchert, 2020). SP is a specific sensitivity to particular situational characteristics (Lawrence, 2006), whereas trait anger is a broader concept comprising of both the disposition to feel angry *without* situational triggers and angry responses to specific (e.g., frustrating or provoking) situations (Spielberger &

Reheiser, 2009). It is possible that sensitivity to particular situational cues might be a key factor in hostile attributions. When it comes to ambiguous situations, typically people focus longer on non-hostile cues as compared to hostile (Wilkowski et al., 2007), but high SP individuals are less sensitive to non-hostile cues, and it is possible that the perception of anger in others might be more habitual for them (Zajenkowska & Rajchert, 2020). This propensity, in turn, may lead to aggressive responses, especially when provoked (Lawrence & Hutchinson, 2013).

The current study

The current study investigates whether sensitivity to provocation predicts bias towards the interpretation of ambiguous facial cues as angry and whether this is depended on the participants' gender and target's sex as well as the participant's tendency to violent behaviour (demonstrated by their imprisonment for violent crimes). We predict that higher sensitivity to provocation will be related to an increased number of anger interpretations (hostility bias) in ambiguous morphed faces (H1). Additionally, we assume that this effect will be stronger in men (H2) as men show greater responses to threatening cues and especially in case of male targets (H3), as evidence has shown that men are superior to females in recognizing male anger (Kret & De Gelder, 2012; Rotter & Rotter, 1988). Also, we stipulate that any gender difference in the bias towards detecting anger in morphed faces between those sensitive to provocation, will be reduced among violent inmates (H4). That is because the more general propensity to interpret facial expressions as angry and hostile among individuals prone to aggressive behavior (including those imprisoned for violent crimes as in the present study) (Mellentin et al., 2015) may over-ride the gender effects.

Also considering that no gender differences have been found in sensitivity to provocations among inmates (Zajenkowska et al., 2013), although they have been observed in general populations of young adults (Zajenkowska et al., 2014), it is possible that a women inmates may demonstrate a threshold to recognize hostile cues similar to that of men inmates.

Method

Note on data

The study was part of a larger research project, for which underlying data can be accessed at www.apsycholab.pl. All analyses and results reported here are novel.

Participants

The sample (N = 283) consisted of 116 prison inmates and 167 adults living in the community. From the initial sample 11 participants' data were excluded from the analysis due to missing data in age, SP or more than 10% missing answers in the morph task. The final sample (N = 272, 138 women[1]) consisted of 105 prison inmates and 167 adults living in the community (aged: $M = 33.46$, $SD = 11.42$).

Prisoners: The final sample of inmates consisted of 105 participants with average age 37.13 ($SD = 9.54$), with 53 women (aged: $M = 37.13$, $SD = 8.81$) and 52 men (aged: $M = 37.13$, $SD = 10.32$). The difference in age between women and men inmates was not statistically significant, $t(99.93) = -0.001$, $p = 0.999$. We informed the prison authorities that individuals eligible to take part were only violent offenders i.e., only those with "any criminal charge for a violent offence against persons – e.g., assault, assault causing bodily harm, wounding, attempted homicide, homicide, kidnapping, forcible confinement, armed robbery, and all 'hands-on' general offences" (Harris et al., 2002, p. 383). The offender groups were recruited from closed wings of two cooperating correctional facilities and according to the report from the prisons' authorities. Information about the study was distributed only among violent inmates. Those willing to participate were provided with an overview of the study by the prison employees. If they chose to participate, the research procedure was initiated and an individual meeting with an experimenter was set.

Community adults: the final non-forensic community sample consisted of 167 participants with no history of previous conviction with an average age of 31.15 years ($SD = 11.91$). There were 85 women (aged: $M = 30.67$, $SD = 13.11$) and 82 men (aged: $M = 31.65$, $SD = 10.57$). The difference in age between women and men was not statistically significant, $t(160.01) = -0.53$, $p = 0.60$. The participants from the community-dwelling sample were recruited in attempt to best match the inmates in terms of age, gender and education. Therefore, we used purposive sampling with predefined sociodemographic criteria and deliberately chose participants due to these characteristics (Etikan et al., 2016). They were recruited via social media or asked directly by investigators in their communities. Nonetheless, on average, inmates were significantly older than non-inmates $t(254.48) = 4.57$, $p < 0.001$. Prisoners also significantly differed from non-inmates on education level, $\chi^2(4) = 58.03$,

[1] In the current study we asked participants to report their *płeć* (in Poland were the study was conducted, this term can be understood both as biological sex or gender). Participants had an open response to this question, and could self-define. All defined as either man or women.

$p < 0.001$. Most of the inmates (42%) had primary education while most of the non-inmates (45%) had secondary education as their highest level completed.

Measures

Sensitivity to Provocation (SP) was measured using the Situational Triggers of Aggressive Responses (STAR) scale (Lawrence, 2006). The full questionnaire consists of 22 items (10 concerning Sensitivity to Frustration - SF, and 12 concerning Sensitivity to Provocation - SP). In present study, only the SP subscale was used. Participants provided a rating of how much they agreed that each of 12 situations would cause them aggressive feelings using a 5-point, Likert scale. Example items included: *A friend betrays me; I am the subject of a practical joke; Someone insults me; Others around me are becoming aggressive; Someone makes offensive remarks to me; I argue with a friend.* This instrument has high internal consistency ($\alpha = 0.82$ for SP) and demonstrated validity in previous studies (Lawrence, 2006). Cronbach's alpha in the present study was $\alpha = 0.90$ (inmates $\alpha = 0.93$, controls $\alpha = 0.88$).

Hostility Bias was measured by a computer task, which assessed the perception of emotional expressions. The task presented a series of morphed faces that were used in previous studies (Schönenberg & Jusyte, 2014) and provided by the authors. Face morphing was achieved by mixing three emotional expressions (angry, happy and fearful) using a computer program (FantaMorph software, Abrosoft, Beijing, China) to create continuous dimensions (happy–angry, fearful–angry, happy-fearful). For each dimension, five intensity levels containing different amounts of each of the blended emotions were established. For instance, a stimulus sequence for the happy–angry dimension contained the following proportions of these emotions in the morphs created: 90% angry and 10% happy; 70% angry and 30% happy; 50% angry and 50% happy (maximal ambiguity); 30% angry and 70% happy; and 10% angry and 90% happy. The stimulus material for this experiment comprised 90 images (faces of three different models; three types of emotions merged and five intensity levels) for both male and female faces. A different set of stimuli was employed for the practice trials. The experimental task consisted of a total of 180 trials (because there were two repetitions), which were randomized across emotions and intensity ratios (Schönenberg & Jusyte, 2014). Female faces were blocked together and the same was done for male faces, and the order was randomized. Each trial started with the presentation of a fixation cross between 1000-2000ms. Subsequently, each morph was displayed between 500-2000ms and participants were asked to choose which emotion was expressed by the faces: angry, fearful or happy. Across the trials, the stimulus presentation time varied to reduce the influence of anticipation on simultaneously recorded eye movement data. Ambiguous morphed faces were those with 50% angry / 50% happy, 50% angry / 50% fearful; 30% angry / 70% happy, 30% angry / 70% fearful and 70% angry / 30% happy, 70% angry / 30% fearful.

Highest education level was reported by the respondents by indicating one of the given grades: primary, secondary, high school, higher, other.

Procedure

The study was conducted by trained senior psychology students from the research team, who also recruited the non-inmate control participants (e.g., through online advertisements). Data were collected in a designated university laboratory or designated rooms within the prisons (e.g., a common room). The participants completed the tasks for the current study as part of a larger research project. The full procedure was as follows: first, the participants read the instructions in which they were informed that they could discontinue at any time. Participants were also informed that their individual data would be handled with confidentiality and would be revealed only as part of an anonymized data set. Subsequently, the participants completed the hostility bias emotion recognition task (some participants also participated at this point in a visual scene assessment task – the results of which are not reported here). Participants also completed paper-pencil questionnaires and the order was counter-balanced meaning that approx. 50% of participants started with the questionnaires regarding personality assessment (e.g. self-description) followed by the computer task/s. All participants ended the study by completing the questionnaires assessing trait anger and sensitivity to provocation. Upon completion of the computer task/s and the questionnaires, participants were debriefed and thanked for their participation.

Results

Data analysis

Statistical analyses desig

The statistical analyses were conducted using the statistical programming language R (R Core Team, 2019). We implemented three Linear Mixed Models (LMM) analysis with crossed random effects for target emotion intensity (Bryk & Raudenbush, 1987), in order to test the moderating role of participant's gender and target's sex on the relationship between sensitivity to provocation (SP) and anger perception in morphed faces, for inmates and non-inmates. First, an analysis was conducted with perception of anger (number

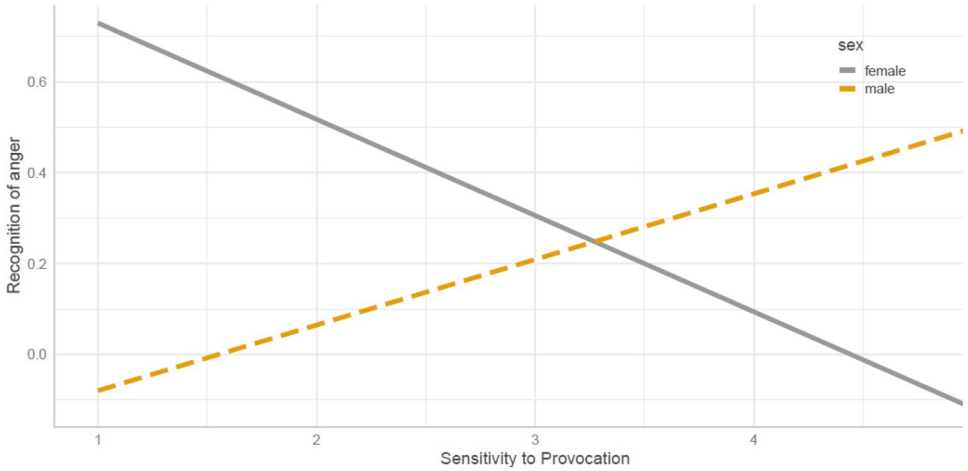


Fig. 1 Sex moderating the slope of sensitivity to provocation on anger recognition on ambiguous faces morphing between anger and happy
Y-axis represents estimated number of recognitions of anger in ambiguous (angry-fear) stimuli, X-axis represents the level of sensitivity to provocation measured on scale from 1 to 5

of anger recognitions, that is hostility bias) as the dependent variable on ambiguous stimuli scenes (emotions morphed with following intensities: 30%/70%, 50%/50%, 70%/30%) differentiating between anger and happiness. The second analysis used perception of anger (number of anger recognitions, that is hostility bias) as the dependent variable, on ambiguous scenes (emotions morphed with following intensities: 30%/70%, 50%/50%, 70%/30%) differentiating between anger and fear. In the third analysis, perception of fear (number of fear recognitions), was the dependent variable on ambiguous scenes (emotions morphed with following intensities: 30%/70%, 50%/50%, 70%/30%) between fear and happiness.

All fitted models used the same fixed effects structure. SP, participants' gender (man vs. women), group (inmate vs. non-inmate), participants' age[2], education level, and target's sex (male vs. female) were treated as fixed factors.

Following the hypotheses, the following interaction terms were also tested in the models: (i) SP x participants' gender, (ii) SP x group, (iii) SP x targets' sex, (iv) SP x participants' gender x targets' sex, (v) participants' gender x group, (vi) participants' gender x targets' sex, (vii) SP x participants' gender x group, (viii) SP x participants' gender x targets' sex, and (ix) SP x participants' gender x targets' sex x group.

In all following LMM analyses, we fitted the base model with random slopes and intensity of anger as a fixed effect, then the full model with all other fixed effects and interaction terms. Full model fixed effects parameters and their p-values were estimated with Satterthwaite's adjustment. Finally, statistically significant interaction effects were followed by linear trends analysis.

[2] Age may impact emotion ratings of faces (Mather & Carstensen, 2005) and education is related to age, as a result, age was added into the model.

In order to test whether male and female prisoners and non-prisoners differ in terms of sensitivity to provocation, a two-way between-subject analysis of variance was conducted with group (prisoners vs. non-prisoners) and gender (men vs. women) as the between-subject variables. Sensitivity to provocation (SP) was treated as a dependent variable. There was a significant effect of sex, $F(1, 272) = 7.96$, $p = 0.005$, $\eta 2p = 0.029$ (women $M = 3.72$, $SD = 0.85$, men $M = 3.44$, $SD = 0.84$). Interestingly, neither the difference in SP between inmates and non-non-inmates, $F(1, 272) < 1$ (inmates $M = 3.56$, $SD = 0.97$, non-inmates $M = 3.59$, $SD = 0.78$), nor the interaction effect of gender and group reached the statistical significance level, $F(1, 272) = 1.40$.

Perception of anger on angry/happy (A/H) faces

Although the main effect of participants' gender was not significant ($F(1,259) = 2.37$, $p = 0.12$), the interaction between participants' gender and SP was marginally significant ($F(1,259) = 3.36$, $p = 0.07$). The follow up linear trends analysis revealed a negative slope between SP and women's anger perception ($\beta = -0.09$, $SE = 0.08$) and a positive slope for men's anger perception ($\beta = 0.12$, $SE = 0.08$), see Fig. 1. As expected, the difference between these slopes was marginally significant ($t(259) = 1{,}83$, $p = 0.068$).

Perception of anger on angry/fearful (A/F) faces

There was a significant main effect of age ($F(1,259) = 8.57$, $p = 0.003$). The age - anger perception slope was significant and positive ($\beta = 0.01$, $SE = 0.004$, $t(259) = 2.93$, $p = 0.004$).

The analysis revealed a main effect of participants' gender ($F(1,259) = 5.75$, $p = 0.017$). The estimated means for the gender effect demonstrated that, on average, men perceived less anger ($M = 4.88$, $SE = 1.63$) than women ($M = 4.93$, $SE = 1.62$) but this effect needs to be considered with great caution as further we had significant interactive

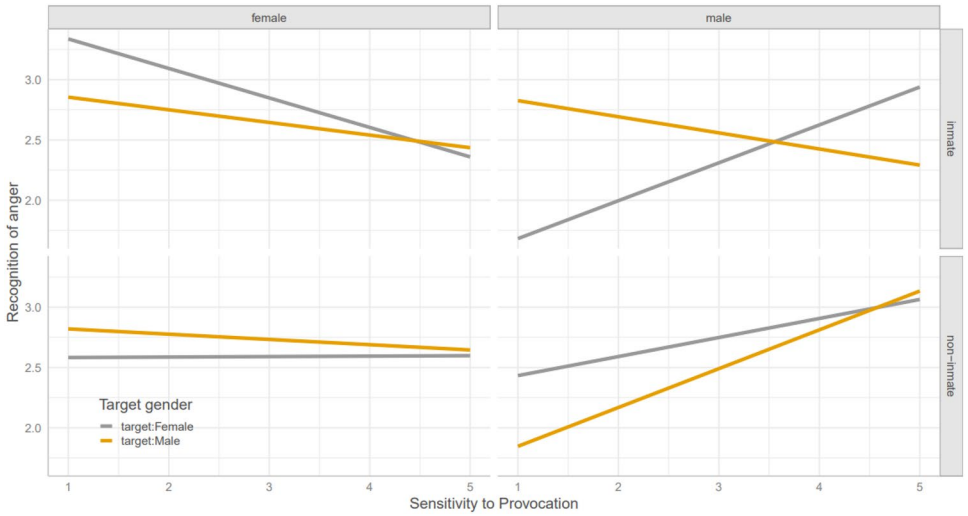


Fig. 2 Interaction effect of SP x participants' gender x target's sex x group on anger recognition on ambivalent faces expressing anger & fear **Y-axis** represents estimated number of recognitions of anger in ambiguous (angry-happy) stimuli, X-axis represents the level of sensitivity to provocation measured on scale from 1 to 5

effects. Although there was no significant main effect of SP ($F(1,259) = 0.39$, $p = 0.532$), the interaction between participants' gender and SP was significant ($F(1, 259) = 5.83$, $p = 0.016$), the interaction between participants' gender, group and targets' sex was significant ($F(1, 1350) = 8.74$, $p = 0.003$), and further, the 4-way interaction between SP, participants' gender, group and targets' sex was significant ($F(1, 1350) = 7.21$, $p = 0.007$). The latter 4-way interaction decomposed using simple trends analysis showed that in the inmate group, the higher SP in men, the higher the identification of anger on female target faces ($\beta = 0.32$, $SE = 0.12$) but the lower on male faces ($\beta = -0.13$, $SE = 0.12$). The difference between these two slopes was statistically significant ($t(1350) = 3.25$, $p = 0.03$). The opposite pattern was observed for women in the inmate group. The higher SP in women, the lower the identification of anger on female target faces ($\beta = -0.24$, $SE = 0.14$). However these women did not show higher identification of anger on male target faces ($\beta = -0.10$, $SE = 0.14$). The difference between these two slopes was not statistically significant ($t(533) = 3.25$, $p = 0.50$), see Fig. 2.

The simple slopes analysis also revealed the marginally significant ($t(539) = 2.98$, $p = 0.06$) difference between slopes for inmate and non-inmate women on recognizing anger on female targets. Examining the women's data showed that for inmates the slope between SP and identification of anger was negative ($\beta = -0.244$, $SE = 0.14$) while for non-inmates the slope was slightly positive ($\beta = 0.003$, $SE = 0.12$), see Fig. 2.

Additionally we observed that for community-based men, SP predicted significantly anger recognition on male faces ($\beta = 0.032$, $SE = 0.14$), see Fig. 2.

Perception of fear on fearful/happy (A/F) faces

There were no significant main effects of participants' gender ($F(1, 259) < 1$) nor SP ($F(1, 259) < 1$). However, the analysis revealed a marginally significant main effect of group ($F(1, 259) = 3.07$, $p = 0.081$), with inmates perceiving fear slightly more ($M = 2.84$, $SE = 1.16$) than the non-inmate group ($M = 2.78$, $SE = 1.15$).

Also, the interaction between SP and group revealed a marginally significant effect ($F(1, 259) = 3.82$, $p = 0.052$). The following linear trends analysis showed that the SP - fear perception slope was positive for the inmate group ($\beta = 0.10$, $SE = 0.08$) and negative for the non-inmate group ($\beta = -0.12$, $SE = 0.08$), meaning that for community adults, as sensitivity to provocation increased, fear perception decreased, see Fig. 3. The direct comparison of slopes reached marginal significance ($t(259) = 1.96$, $p = 0.052$).

Discussion

The current study investigated whether sensitivity to provocation was linked to a bias towards perceiving anger in human faces among women and men convicted for violent crimes compared to men and women drawn from the community. In particular, we predicted that higher sensitivity to provocation would be related to an increased number of anger interpretations (hostility bias) of ambiguous morphed faces (H1). We assumed that this effect would be stronger in men (H2) and especially in case of male targets (H3). Finally, we expected that gender difference in the bias towards detecting anger in morphed faces between those sensitive to provocation, would be reduced among violent inmates (H4).

Our hypotheses were confirmed partially, as hostility bias was related differently to sensitivity to provocation across genders and groups depending on the target's sex. Generally, in the inmate group, the higher the sensitivity to provocation in men, the higher identification of anger in female target faces, but the lower the identification in male faces (anger/

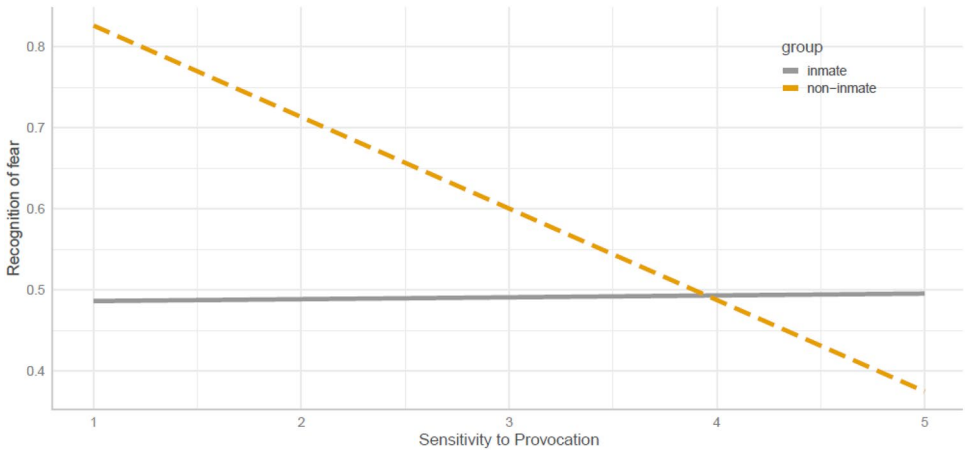


Fig. 3 Group moderating the slope of sensitivity to provocation on fear recognition on ambiguous faces morphing between fear and happy **Y-axis** represents estimated number of recognitions of fear in ambiguous (fear - happy) stimuli, X-axis represents the level of sensitivity to provocation measured on scale from 1 to 5

fear morphs). On the contrary, the higher the sensitivity to provocation in women, the lower the identification of anger in female target faces (anger/fear morphs). However, women did not show the tendency to better identify anger in male target faces. Additionally, in the case of non-inmates, men who were higher in SP identified anger in male faces to a greater extent.

In the case of anger/happy morphs, women who were more sensitive to provocation, perceived less anger in ambiguous faces. For men who were sensitive to provocation, the opposite was the case: they perceived anger more. However, it should be noted that these latter effects were only marginally significant. Gender differences were visible mostly in faces which morphed angry and fearful expressions, but there was also the same trend displayed in morphs of angry and happy expressions. Moreover, gender differences in a hostility bias, and not in a general bias toward negative affect were demonstrated as these were not observed in case of fear/happy morphs. At the same time there were differences between inmates and non-inmates prone to sensitivity to provocation and fear recognition.

The findings that men who are sensitive to provocation perceive anger in faces more readily is predictable. Such gender differences in hostile attributions may be a function of evolutionary patterns in the risks for, and benefits from, engaging in social encounters, as men perceive more adversity in neutral settings compared to women (Jonason et al., 2021). Aggressive responses to provocation are influenced by cultural and social norms dictating what is perceived 'appropriate' for men and women to do, and in what contexts (Thomas & Atakan, 1993). Anger is often associated with attack (e.g., Berkowitz 1993) and is an approach-oriented emotion, which may result in a tendency to aggress (Harmon-Jones et al., 2011; Lewis, 1993). Although there is a general agreement that there is no difference in trait anger between men and women, both genders can differently express this emotion, which may be related to what men and women are socially "sanctioned" to do in certain social and cultural contexts (Thomas & Atakan, 1993). This is potentially the case in the current study, as it was conducted in a relatively non-egalitarian culture: Poland (Zawisza et al., 2015), with clearly established gender roles. It is plausible that in this context, women typically try to 're-frame' angry or provoking stimuli and feelings in order to conform to a non-aggressive gender role. Further work should examine participants' own traditional or egalitarian gender identity to confirm whether this does indeed moderate the link between sensitivity to provocation and anger detection for women.

However, it could be also that women sensitive to provocation might look for situations, which guarantee them being in a position of power. Perceiving another person as frightened, i.e. detecting fear, could be a factor that also leads to aggression. For instance, girls with conduct problems and elevated callous-unemotional traits outperform girls with no conduct problems in the identification of fear (Schwenck et al., 2014). The opposite is true for comparable boys, as boys with psychopathic tendencies (Blair et al., 2001) and male psychopaths (Blair et al., 2001) showed reduced fear recognition, and in general there is an agreement that psychopathic traits are related to lower and less accurate fear recognition (Marsh & Blair, 2008). It is possible that different emotion recognition biases may lead to aggressive behaviors among men and women prone to provocation. As a result, future studies could investigate how anger and fear recognition among men and women relate to aggressive emotional responses and the inclination to aggress. That is especially vital as among inmates men sensitive to provocation recognizes anger more often on female faces and less on male faces (that is more fear). Imprisoned women recognized less anger (more fear) on female faces. It is possible that seeing anger on a face of person who is perceived as physically weaker may be an impulse or even a justification to attack.

Supplementary when interpreting anger recognition on female faces, we could refer to object relation theory and early representation of relation with a primary caregiver,

which usually is a women (McWilliams, 2011). Inmates quite often come from disfunction families, and, indeed, insecure attachment style is linked to psychopathic behavior among prison inmates (Onyedibe, 2011). It is possible that the female targets presented during the study awaken different feelings in men and women prone to violence due to differences in the ability to express and process anger in relation to their own mothers. Ideally, future studies examining the perception of emotion in others' faces should investigate their relationships with significant caregivers (e.g. qualitative analysis) and participants' attachment styles.

Hostility bias, assessed here as the tendency to see anger in faces, is one of the examples of emotional biases in attention and there are different underlying tendencies associated with perceiving emotional expressions in a maladaptive and consistent manner (Mellentin et al., 2015). For individuals who are particularly sensitive to provocation, it is possible that the perception of specific emotions would lead to aggression, depending on their gender or the group they belong to. These possibilities need further investigation, but it could lead to useful clinical interventions. Novel clinical neuropsychological methods have targeted emotional dysfunctions, e.g. using Attentional Bias Modification (ABM) by correcting emotional biases (e.g. Beard et al., 2012). However, it is vital, first, to examine and verify, which emotion recognition leads to what behavior. For example, it is possible that for some, perceiving anger, and for others perceiving fear or happiness, would lead to hostile interpretations and, in turn, aggression.

Limitations

A notable limitation of the current study is the fact that we assessed sensitivity to provocation using self-report questionnaire. It would a useful alternative to use other assessment methods, including behavioral (e.g. methods using adaptations of Taylor's (1967) Reaction Time Tasks) or physiological measures (e.g. skin conductance; Seidl et al., 2013) to assess sensitivity to the triggers of provocation and hostile arousal for both male and female offenders and non-offenders. Moreover, all participants, as well as the targets portrayed the emotional expressions, were white/Caucasian so it should be noted that these findings may not generalize to other-race stimuli. Also, in our study we used non-naturalistic (morphed) facial expressions, which limits the ability to draw inferences for real life behavior. This could be addressed in future studies by introducing real life interactions with a partner expressing different emotions. Additionally, there were differences between inmates and non-inmates in terms of age and highest level of education. This could influence the findings as increased anger recognition from faces with morphed anger and fear is related to increased age (Mather & Carstensen, 2005). Therefore, although it is a challenge, greater care is needed to match the samples in terms of these relevant socio-demographic factors.

Finally, we were not able to collect information about inmates who did not opt to participate in the study, which might, of course, have had an impact on the results. Future studies would ideally collect basic demographic information from those who did not participate in order to exclude alternate explanations for findings.

To sum up, the current project can be considered a step forward in examining gender differences in hostile intent of ambiguous stimuli in offenders and community samples. Such studies are important because little work in this field has been conducted in female offending samples and fewer still examine gender differences in factors related with hostile intent among offenders and community samples.

Funding The studies were founded by National Science Centre (UMO-2017/26/D/HS6/00258).

Data availability Data used in this study will be stored at: http://www.apsycholab.pl.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures followed were in accordance with the ethical standards of the responsible. Committee on human experimentation of Maria Grzegorzewska University (acceptance of the Ethics Committee) and with the Helsinki Declaration. Informed consent was obtained from all participant included in the study.

References

Ayduk, Ö., Gyurak, A., & Luerssen, A. (2008). Individual differences in the rejection–aggression link in the hot sauce paradigm: The case of rejection sensitivity. *Journal of Experimental Social Psychology, 44*(3), 775–782.

Archer, J. (2004). Gender differences in aggression in real world settings: A meta-analytic review. *Review of General Psychology*, *8*, 291–322

Austin, J. B., Bloom, B. E., & Donahue, T. (1993). *Female offenders in the community: An analysis of innovative strategies and programs*. US Department of Justice, Office of Justice Programs, National Institute of Justice

Beard, C., Sawyer, A. T., & Hofmann, S. G. (2012). Efficacy of attention bias modification using threat and appetitive stimuli: A meta-analytic review. *Behavior Therapy, 43*(4), 724–740.

Bennett, S., Farrington, D. P., & Huesmann, L. R. (2005). Explaining gender differences in crime and violence: The importance of social cognitive skills. *Aggression and Violent Behavior, 10*(3), 263–288.

Berkowitz, L. (1993). *Aggression: Its causes, consequences, and control*. New York: Mcgraw-Hill Book Company.

Blair, R. J. R. (2003). Neurobiological basis of psychopathy. *The British Journal of Psychiatry*, *182*(1), 5–7

Blair, R. J. R., Colledge, E., Murray, L., & Mitchell, D. G. V. (2001). A selective impairment in the processing of sad and fearful expressions in children with psychopathic tendencies. *Journal of Abnormal Child Psychology, 29*(6), 491–498.

Blair, R. J. R., Mitchell, D. G. V., Peschardt, K. S., Colledge, E., Leonard, R. A., Shine, J. H. … Perrett, D. I. (2004). Reduced sensitivity to others' fearful expressions in psychopathic individuals. *Personality and Individual Differences*, *37*(6), 1111–1122

Bondü, R., & Richter, P. (2016). Interrelations of justice, rejection, provocation, and moral disgust sensitivity and their links with the hostile attribution bias, trait anger, and aggression. *Frontiers in Psychology, 7*, 795.

Brazão, N., Motta, D., Rijo, C., Céu, D., Salvador, M., Pinto-Gouveia, J., & Ramos, J. (2015). Clinical change in cognitive distortions and core schemas after a cognitive–behavioral group intervention: Preliminary findings from a randomized trial with male prison inmates. *Cognitive Therapy and Research, 39*(5), 578–589.

Brazão, N., Motta, D., Rijo, C., Céu, D., Salvador, M., Pinto-Gouveia, J., & Ramos, J. (2015). Clinical change in cognitive distortions and core schemas after a cognitive–behavioral group intervention: Preliminary findings from a randomized trial with male prison inmates. *Cognitive Therapy and Research*, *39*(5), 578–589

Bushman, B. J. (2016). Violent media and hostile appraisals: A meta-analytic review. *Aggressive Behavior, 42*(6), 605–613.

Carlson, C. R., Felleman, E. S., & Masters, J. C. (1983). Influence of children's emotion states on the recognition of emotion in peers and social motives to change another's emotional state. *Motivation and Emotion*, *7*, 61–79

Combs, D. R., Penn, D. L., Wicher, M., & Waldheter, E. (2007). The ambiguous intentions hostility questionnaire (AIHQ): A new measure for evaluating hostile social cognitive biases in paranoia. *Cognitive Neuropsychiatry, 12*(2), 128–143.

Corden, B., Critchley, H. D., Skuse, D., & Dolan, R. J. (2006). Fear recognition ability predicts differences in social cognitive and neural functioning in men. *Journal of Cognitive Neuroscience, 18*(6), 889–897.

Deffenbacher, J. L. (1992). Trait anger: Theory, findings, and implications. In C. D. Spielberger & J. N. Butcher (Eds.), *Advances in personality assessessment* (Vol. 9, pp. 177–201). Hillsdale, NJ: Erlbaum

Dodge, K. A. (2006). Translational science in action: Hostile attributional style and the development of aggressive behavior problems. *Development and Psychopathology, 18*(3), 791–814.

Dodge, K. A., Price, J. M., Bachorowski, J. A., & Newman, J. P. (1990). Hostile attributional biases in severely aggressive adolescents. *Journal of abnormal psychology*, *99*(4), 385

Downey, G., Lebolt, A., Rincón, C., & Freitas, A. L. (1998). Rejection sensitivity and children's interpersonal difficulties. *Child Development, 69*(4), 1074–1091.

Driscoll, H., Zinkivskay, A., Evans, K., & Campbell, A. (2006). Gender differences in social representations of aggression: The phenomenological experience of differences in inhibitory control? *British Journal of Psychology*, *97*(2), 139–153

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics, 5*(1), 1–4.

Fridlund, A. J. (1991). Evolution and facial action in reflex, social motive, and paralanguage. *Biological Psychology, 32*(1), 3–100.

Harmon-Jones, C., Schmeichel, B. J., Mennitt, E., & Harmon-Jones, E. (2011). The expression of determination: Similarities between anger and approach-related positive affect. *Journal of Personality and Social Psychology*, *100*(1), 172

Harris, G. T., Rice, M. E., & Cormier, C. A. (2002). Prospective replication of the Violence Risk Appraisal Guide in predicting violent recidivism among forensic patients. *Law and Human Behavior*, *26*(4), 377–394

Horsley, T. A., de Castro, B. O., & Van der Schoot, M. (2010). In the eye of the beholder: Eye-tracking assessment of social information processing in aggressive behavior. *Journal of abnormal child psychology*, *38*(5), 587–599

Jonason, P. K., Zajenkowski, M., & Leniarska, M. (2021). From situational perceptions to personality pathologies. *Personality and Individual Differences*, *181*, 111049

Knight, G. P., Guthrie, I. K., Page, M. C., & Fabes, R. A. (2002). Emotional arousal and gender differences in aggression: A meta-analysis. *Aggressive Behavior*, *28*, 366–393

Kret, M. E., & De Gelder, B. (2012). A review on sex differences in processing emotional signals. *Neuropsychologia*, *50*(7), 1211–1221

Kret, M. E. (2011). Context, gender and personality factors, influencing the perception of facial and bodily expressions of emotion. Doctoral dissertation. Ridderkerk: Ridderprint

Lawrence, C. (2006). Measuring individual responses to aggression-triggering events: Development of the situational triggers of aggressive responses (STAR) scale. *Aggressive Behavior*, *32*(3), 241–252

Lawrence, C., & Hodgkins, E. (2009). Personality influences on interpretations of aggressive behavior: The role of provocation sensitivity and trait aggression. *Personality and Individual Differences*, *46*(3), 319–324

Lawrence, C., & Hutchinson, L. (2013). The influence of individual differences in sensitivity to provocations on provoked aggression. *Aggressive Behavior*, *39*(3), 212–221

Lewis, J. E. (1993). *Women and social policies in Europe: Work, family and the state*. Edward: Edward Elgar Pub.

Logan, C., & Blackburn, R. (2009). Mental disorder in violent women in secure settings: Potential relevance to risk for future violence. *International Journal of Law and Psychiatry*, *32*, 31–38

Mandal, M. K., & Palchoudhury, S. (1985). Perceptual skill in decoding facial affect. *Perceptual and Motor Skills*, *60*(1), 96–98

Marsh, A. A., & Blair, R. J. R. (2008). Deficits in facial affect recognition among antisocial populations: A meta-analysis. *Neuroscience & Biobehavioral Reviews*, *32*(3), 454–465

McWilliams, N. (2011). *Psychoanalytic diagnosis: Understanding personality structure in the clinical process*. New York: Guilford Press.

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

Montagne, B., van Honk, J., Kessels, R. P. C., Frigerio, E., Burt, M., van Zandvoort, M. J. E., & de Haan, E. H. F. (2005). Reduced efficiency in recognising fear in subjects scoring high on psychopathic personality characteristics. *Personality and Individual Differences*, *38*(1), 5–11

Muller, E., & Kempes, M. (2016). Gender differences in a Dutch forensic sample of severe violent offenders. *International journal of forensic mental health*, *15*(2), 164–173

Ogle, R. S., Maier-Katkin, D., & Bernard, T. J. (1995). A theory of homicidal behavior among women. *Criminology*, *33*(2), 173–193

Onyedibe, M. C. C. (2011). Attachment style and self-esteem as Predictors of Psychopathic behaviour among male Prison inmates. *Nigerian Journal of Psychological Research, 7*, 25–41.

Quigley, B. M., & Tedeschi, J. T. (1996). Mediating effects of blame attributions on feelings of anger. *Personality and Social Psychology Bulletin*, *22*(12), 1280–1288

R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Rotter, N. G., & Rotter, G. S. (1988). Sex differences in encoding and decoding of negative facial emotion. *Journal of Nonverbal Behavior*, *12*, 139–148

Schiffenbauer, A. (1974). Effect of observer's emotional state on judgments of the emotional state of others. *Journal of Personality and Social Psychology*, *30*(1), 31

Schönenberg, M., & Jusyte, A. (2014). Investigation of the hostile attribution bias toward ambiguous facial cues in antisocial violent offenders. *European Archives of Psychiatry and Clinical Neuroscience*, *264*(1), 61–69

Schönenberg, M., Jusyte, A., Höhnle, N., Mayer, S. V., Weber, Y., Hautzinger, M., & Schell, C. (2015). Theory of mind abilities in patients with psychogenic nonepileptic seizures. *Epilepsy & Behavior*, *53*, 20–24

Schönenberg, M., Mayer, S. V., Christian, S., Louis, K., & Jusyte, A. (2016). Facial affect recognition in violent and nonviolent antisocial behavior subtypes. *Journal of Personality Disorders*, *30*(5), 708–719

Schultz, D., Grodack, A., & Izard, C. E. (2010). *State and trait anger, fear, and social information processing. In International handbook of anger* (pp. 311–325). New York, NY: Springer

Schwenck, C., Gensthaler, A., Romanos, M., Freitag, C. M., Schneider, W., & Taurines, R. (2014). Emotion recognition in girls with conduct problems. *European child & adolescent psychiatry*, *23*(1), 13–22

Seidl, E. M., Pfabigan, D. M., Keckeis, K., Wucherer, A. M., Jahn, T., Lamm, C., & Derntl, B. (2013). Empathic competencies in violent offenders. *Psychiatry Research*, *210*, 1168–1175

Shonkoff, J. P., Phillips, D. A., & National Research Council. (2000). Making friends and getting along with peers. *From Neurons to Neighborhoods: The Science of Early Childhood Development*. National Academies Press (US)

Snell, T. L., & Morton, D. C. (1994). *Women in prison*. Washington, DC: US Department of Justice, Office of Justice Programs, National Institute of Justice

Spielberger, C. D., & Reheiser, E. C. (2009). Assessment of emotions: Anxiety, anger, depression, and curiosity. *Applied Psychology: Health and Well-Being*, *1*(3), 271–302

Taylor, S. (1967). Aggressive behavior and physiological arousal as a function of provocation and the tendency to inhibit aggression. *Journal of Personality*, *35*, 297–310

Thomas, S. P., & Atakan, S. (1993). Trait anger, anger expression, stress, and health status of American and Turkish midlife women. *Health Care for Women International*, *14*(2), 129–143

Tuente, S. K., Bogaerts, S., & Veling, W. (2019). Hostile attribution bias and aggression in adults-a systematic review. *Aggression and Violent Behavior, 46*, 66–81.

Veague, H. B., & Hooley, J. M. (2014). Enhanced sensitivity and response bias for male anger in women with borderline personality disorder. *Psychiatry Research*, *215*(3), 687–693

Verona, E., & Carbonell, J. L. (2000). Female violence and personality: Evidence for a pattern of overcontrolled hostility among one-time violent female offenders. *Criminal Justice and behavior*, *27*(2), 176–195

Wagner, A. W., & Linehan, M. M. (1999). Facial expression recognition ability among women with borderline personality disorder: implications for emotion regulation? *Journal of Personality Disorders, 13*(4), 329–344.

Wilkowski, B. M., Robinson, M. D., Gordon, R. D., & Troop-Gordon, W. (2007). Tracking the evil eye: Trait anger and selective attention within ambiguously hostile scenes. *Journal of Research in Personality*, *41*(3), 650–666

Zajenkowska, A., Jankowski, K. S., Lawrence, C., & Zajenkowski, M. (2013). Personality and individual differences in responses to aggression triggering events among prisoners and non-prisoners. *Personality and Individual Differences, 55*(8), 947–951.

Zajenkowska, A., Mylonas, K., Lawrence, C., Konopka, K., & Rajchert, J. (2014). Cross-cultural gender differences in situational triggers of aggressive responses. *International Journal of Psychology*, *49*(5), 355–363

Zajenkowska, A., Prusik, M., Jasielska, D., & Szulawski, M. (2021). Hostile attribution bias among offenders and non-offenders: Making social information processing more adequate. *Journal of Community & Applied Social Psychology, 31*(2), 241–256.

Zajenkowska, A., & Rajchert, J. (2020). How sensitivity to provocation shapes encoding and interpretation of ambivalent scenes in an eye tracking study. *Journal of Cognitive Psychology*, *32*(2), 180–198

Zawisza, M., Luyt, R., & Zawadzka, A. M. (2015). Societies in transition: Are they more genderist? A comparison between Polish, South African and British samples. *Journal of Gender Studies*, *24*, 38–55

Original Article

Hostile Attribution Biases in Vulnerable Narcissists Depends on the Socio-Relational Context

Marta Bodecka-Zych[1], Peter K. Jonason[2,3], and Anna Zajenkowska[1]

[1]Psychology Department, The Maria Grzegorzewska University, Warsaw, Poland
[2]Department of General Psychology, University of Padova, Italy
[3]Institute of Psychology, University of Kardinal Stefan Wyszyński, Warsaw, Poland

Abstract: Narcissism, especially the vulnerable kind, is associated with anger and hostility. In a sample of outpatients, group psychotherapy (N = 74) and community members (N = 153), we replicated and extended previous work linking vulnerable, but not grandiose, narcissism to hostile attributions across different socio-relational contexts. We also examined if the level of ambiguity of social situations, assessed from the other-referent position, influences the relationship between vulnerable narcissism and attributing hostile intentionality, and whether narcissistic individuals distinguish hostile interpretations, depending on the level of ambiguity of the scene. In ambiguous vignettes, assessed from the self-referent position, higher levels of vulnerable narcissism were associated with a greater tendency to infer more attributions of hostile intentions with people with whom there was no close relationship (except for authority in the patient group). In the case of visual scenes, the positive relationship between vulnerable narcissism and attributed hostile intentionality appeared in accidental scenes, but not in hostile and ambiguous ones. In addition, the higher the vulnerable narcissism the lower the ability to differentiate between contextual nuances (e.g., the level of ambiguity). We replicated previous research indicating a relationship between vulnerable narcissism and hostile attribution bias, but shed new light on the phenomenon of this bias in that it appears to depend on the socio-relational context and the level of ambiguity of the situation.

Keywords: vulnerable narcissism, grandiose narcissism, hostile attributions, socio-relational context, psychotherapy group

Narcissistic people are characterized by a sense of superiority, the need to be admired by others, and a lack of empathy (Czarna, 2011). Unsurprisingly, these people tend to also engage in various forms of antisocial behaviors like aggression (Kernberg, 1975; Rasmussen, 2016) and cruelty (Wink, 1991). For instance, they exhibit "narcissistic rage," an explosive mixture of anger and hostility common in ego-threatening situations (Krizan & Johar, 2015) which are often fueled by a disrupted self-image (Kohut, 1972). However, narcissism is not a homogeneous construct, and its various forms are distinguished, for example, by grandiose (aka, overt) and vulnerable (aka, covert) narcissism (Miller et al., 2011, 2013; Pincus & Lukowitsky, 2010; Wink, 1991). Most of the unfavorable and socially undesirable features of narcissism may stem from an inferiority complex in what is called vulnerable narcissism (Miller et al., 2011; Wink, 1991). This form of narcissism is derived from an elevated level of neuroticism, based largely on anxiousness along with threatening and hostile perceptions of reality (Czarna et al., 2019; Dickinson & Pincus, 2003; Miller, et al., 2018). In contrast, grandiose narcissism positively correlates with extraversion, while negatively correlates with neuroticism and agreeableness (Miller et al., 2010, 2011). It is characterized by increased self-esteem, a demanding attitude, and the need to be admired (Dickinson & Pincus, 2003; Wink, 1991). It is also associated with a reduced level of (declared) loneliness or depression, and, therefore, leads to (reported) positive affective states (Sedikides et al., 2004). Both forms of narcissism are associated with aggression; grandiose with proactive and reactive aggression, and vulnerable with reactive (Vize et al., 2019). However, only the second type of narcissism is connected to internal aspects of aggression (Czarna et al., 2019). Unlike grandiose narcissism, vulnerable narcissism appears to be a defensive attitude leading to a heightened sense of hostility from others which often leads to anger (Krizan & Johar, 2015; Wink, 1991). One interesting psychological bias that may be associated with vulnerable narcissism but not grandiose narcissism and potentially at the heart of the psychosocial dysfunctions of people characterized by the former is an over-perception of hostile intentions in others (Hansen-Brown & Freis, 2021). That is, they may

have a hostile attribution bias which leads them to view others and situations as threatening, to react defensively, and engendering rejection and worse from others, all of which create a negative feedback loop. This negative feedback loop may be part of (1) the resistance to treatment characteristic of narcissistic personality disorder (Scheff, 1987; Tangney et al., 1996) and (2) common perceptions around the issues with narcissists (Krizan & Herlache, 2018). Therefore, we predict that vulnerable but not grandiose narcissism will be associated with hostile attribution of intent.

Prior research has confirmed this hypothesis (Hansen-Brown & Freis, 2021) but was limited in several ways. First, as it is common in personality research, the effects have been documented in college students, who experience various types of mental problems (Tosevski et al., 2010), however, clinical populations (e.g., patients) present them to a greater extent (Zajenkowska et al., 2019). Second, there has been an (over)reliance on vignette-based methods that frame the process in a potentially problematic, self-referential way (Jahoda et al., 2006). Furthermore, vignette methods might not factor analysis well as they do not load on a single latent dimension (Zajenkowska et al., 2018). This leads to the third limitation; authors often assume there exists a hostile attributional style akin to a personality trait (Cutrona et al., 1984). Meanwhile, attributions of hostile intent can be sensitive to the socio-relational context (Basquill et al., 2004; Cutrona et al., 1984; Jahoda et al., 2006; Wilkowski et al., 2007; Zajenkowska et al., 2018; Zajenkowska & Rajchert, 2020). This context is differentiated among many by several factors including (1) the relationship described in the stimuli (e.g., friend vs. stranger), (2) the level of provocation/hostility described in the situation (e.g., hostile vs. accidental), and assessment perspective (e.g., self-referent position vs. other-referent position). Therefore, we verify if vulnerable narcissism is associated with a greater tendency to make hostile attributions (i.e., ascribe greater intent and blame to the perpetrator and declare angry feelings) depending on the type of relationship in a particular event and the relative ambivalence of the threats. We do so in samples of clinical patients and community members. Vulnerable narcissism is associated with many clinically relevant criteria of psychopathology, such as interpersonal guilt, maladaptive defense mechanisms, low self-esteem, or reduced psychological well-being (Kaufman et al., 2020). We also test whether the strength of the correlation between vulnerable narcissism and hostile attributions is greater in the clinical sample than in the community sample.

We use two different methods of assessing individual differences in attributed hostile intentions. The first method is based only on ambiguous stimuli (i.e., unclear, which could be interpreted as both hostile and non-hostile). The second method contains hostile, ambiguous, and non-hostile stimuli, which allows us to check whether people with a high level of vulnerable narcissism generally differentiate their causal attributions depending on the ambiguity level of the scene.

A long-held idea regarding narcissism is that they are hostile people (Krizan & Johar, 2015). Underneath such hostility might be a tendency to misread social situations such that narcissistic people see threats most people do not see. However, theoretically, this assertion is limited because narcissism is not always associated with hostility when considering nuanced views of the trait that include the grandiose and vulnerable distinction (Dickinson & Pincus, 2003; Sedikides et al., 2004; Wink, 1991). Therefore, in this study, we attempt to replicate (Hansen-Brown & Freis, 2021) and extend work on how these two traits are related to individual differences in hostile attributions of intent. Importantly, we add some sampling heterogeneity to a field that tends to rely on college students; in Sample 1 we collect data from community members engaged in group psychotherapy.

Method

Participants and Procedure

We collected two samples: A clinical (N = 74; 52 women; $\text{Range}_{\text{Age}}$ = 19–59, M_{Age} = 34.43, SD_{Age} = 8.72) and a non-clinical (N = 153; 97 women; $\text{Range}_{\text{Age}}$ = 18–49, M_{Age} = 25.35, SD_{Age} = 5.50) group of volunteers.[1] Both samples came from a large, Central European city; most of whom from Sample 1 had a higher degree (n = 47) and were single (n = 32) and most of whom from Sample 2 were university students (n = 83) and were single (n = 80). Sample 2 participants were recruited via social media or asked directly by investigators in their community to partake in a larger study in a designated room in a university. Sample 1 involved people in an outpatient, group therapy practice who were seeking therapy for interpersonal, affective or anxiety

[1] We originally collected data from 89 participants but 15 needed to be eliminated for problematic gaps in the data. No participants were excluded from Sample 2. Based on the correlations (r = .25) reported previously (Hansen-Brown & Freis, 2021), we aimed for at least 80% power with a 5% error rate; therefore, we needed to recruit at least 120 participants. This was unfeasible in Sample 1 given the nature of the group. However, in the case of specific samples (e.g., offenders) such a number of participants should still allow us to detect group differences in narcissism because the effects are expected to be larger in clinical groups (e.g., Hepper et al., 2014).

disorders, and personality disorders.[2] Patients were recruited to participate during one-on-one consultations with their therapist, before starting therapy. Individuals with a history of psychosis were excluded a priori. Participants completed the prepared set of questionnaires after the first group session and after the last group session because the study was part of a larger project focused on changes in hostile attributions after group psychotherapy. However, in the present study, we analyzed only answers from the first measurement. Most of the patients completed their questionnaire alone in a group setting where the researcher, not the therapist, was present in the room. In both samples, all participants were assured of the anonymity/confidentiality of their participation, provided informed consent, and were debriefed and thanked for their participation. The study was conducted by the established ethical guidelines and received approval from the hospital authorities.

Measures

In both samples, to measure hostile attributions, we used the Polish translation (Zajenkowska et al., 2018) of the Ambiguous Intentions and Hostility Questionnaire (AIHQ; Combs et al., 2007). Participants were asked to assess whether the other person/s acted on purpose (1 = *definitely no*; 6 = *definitely yes*), how angry it made them feel (1 = *not angry;* 5 = *very angry*), and how much they blamed the other person/s (1 = *not at all*; 5 = *very much*) with five hypothetical, ambiguous situations, that involve a range of social relationships, including a new "co-worker," "an authority figure," "strangers," "an acquaintance," and "a friend." Descriptions of the vignettes are included in the Appendix. The scoring of the AIHQ often consists of summing blame, intentionality, and anger, resulting in one index of hostile attribution across the five scenarios (Combs et al., 2007). This approach reflects the strong inter-correlation between blame, intentionality, and anger (all *r*'s > 70; Combs et al., 2007). However, the AIHQ may include five indices of hostile attributions (one for each of the scenarios, reflecting differences as a function of social relatedness) rather than simply one general hostile attributions dimension, as indicated by previous studies (Zajenkowska et al., 2018). Therefore, we calculated five situational hostile attributions as a mean composite scores of the three questions regarding blame, anger, and intentionality per situation (Coworker $\alpha_{Sample1}$ = .83, $\alpha_{Sample2}$ = .86; Authority $\alpha_{Sample1}$ = .83, $\alpha_{Sample2}$ = .81; Strangers $\alpha_{Sample1}$ = .86, $\alpha_{Sample2}$ = .85; Acquaintance $\alpha_{Sample1}$ = .80, $\alpha_{Sample2}$ = .81; Friend $\alpha_{Sample1}$ = .91, $\alpha_{Sample2}$ = .90).

In Sample 2, in addition to the above method, we also used a scene rating task to measure hostile attributions (Wilkowski et al., 2007; Zajenkowska & Rajchert, 2020), which contained 99 pictures presented on a computer monitor. Participants were asked to rate the extent to which the depicted harm was intentional (1 = *not intended at all*; 9 = *intended*). Each scene was created in three versions: presenting (1) unambiguously hostile behavior between two men/women, (2) unintentional event, and (3) ambiguous event (i.e., some aspects indicated hostile behaviors, and some indicated unintentional ones). In the first version (i.e., unambiguous), all cues suggested hostile behavior of one of the actors (e.g., facial expression and hand/leg direction in the first picture of Figure 1). In the second version, all cues suggested an unintentional situation (e.g., second situation in Figure 1). The ambiguous version included some aspects (e.g., hand/leg direction) indicating hostile behavior and some indicating unintentional behavior (e.g., facial expression: see the third picture in Figure 1). Scenes were prepared about everyday situations and depicted adults only. To create a total score of hostile attribution bias for each type of scene, we averaged the responses separately for Hostile (α = .84), Non-Hostile (α = .87), and Ambiguous (α = .69) scenes.

In both samples, to measure vulnerable narcissism, we used the Polish translation (Czarna et al., 2014) of the 10-item Hypersensitive Narcissism Scale (HSNS; Hendin & Cheek, 1997). Participants were asked for their agreement (1 = *strongly disagree*; 5 = *strongly agree*) with items like "I easily become wrapped up in my own interests and forget the existence of others" and "I am secretly 'put out' when other people come to me with their troubles, asking me for my time and sympathy." In both samples, we averaged participants' ratings to create a score of vulnerable narcissism (Cronbach's $\alpha_{Sample1}$ = .77, *M* = 3.40, *SD* = 0.60; $\alpha_{Sample2}$ = .71, *M* = 2.91, *SD* = 0.55).

In both samples, to measure grandiose narcissism, we used the Polish translation (Banaśkiewicz-Bazińska & Drat-Ruszczak, 2000) of the Narcissistic Personality Inventory (NPI; Raskin & Hall, 1979; Raskin & Terry, 1988) where participants were asked to rate how much (1 = *not all like me*; 5 = *it is highly like me*) 34 items described them (e.g., "I have a natural talent for influencing people" or "I like to show my body").[3] In both samples, we averaged participants' ratings to create a score of grandiose

[2] We wanted to include patients from the same clinic to avoid confounding factors. Because the groups were limited to no more than 12 patients over 3 months, usually no more than two groups were conducted simultaneously.

[3] The Polish version of the NPI questionnaire differs from the original in terms of the number of items (40 in the original) and the response format (dichotomous in the original), the reasons for these changes are described by the authors of the adaptation (Banaśkiewicz-Bazińska & Drat-Ruszczak, 2000).

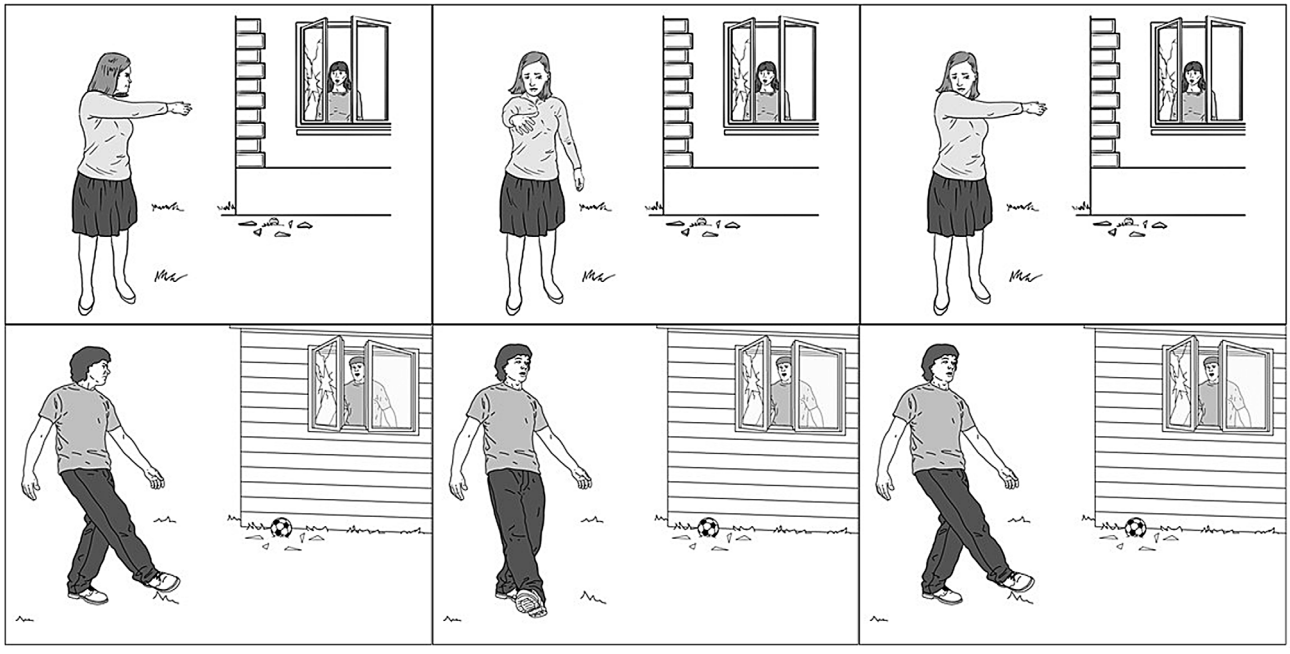


Figure 1. Examples of scenes with male and female actors used in tasks to measure hostile attributions (Wilkowski et al., 2007; Zajenkowska & Rajchert, 2020).

narcissism ($\alpha_{Sample1}$ = .85, *M* = 2.42, *SD* = 0.70; $\alpha_{Sample2}$ = .93, *M* = 3.01, *SD* = 0.59).[4]

Results

All analyses were performed in IBM SPSS 26.0.01 for Windows. To verify the relationship between vulnerable narcissism and the tendency to make hostile attributions, we checked the bivariate correlations, however, we were interested in dependence on the socio-relational context, thus we examined correlations separately for each situation. In Table 1, we present descriptive statistics for the hostile attributions and correlations between those attributions and narcissism. In Sample 1 (top panel), vulnerable narcissism was positively correlated with a hostile attribution bias for an encounter with a coworker, acquaintance, and a stranger only. In Sample 2 (bottom panel), vulnerable narcissism was positively correlated with a hostile attribution bias across all situations except for the one with a friend. The correlations between both types of narcissism and hostile attributions measured with AIHQ were retrieved from two different samples, therefore, we examined them against each other using a Fisher's *z*-test (Eid et al., 2011), revealing no differences in the strength of the correlations between patients and community participants (*z*'s = −0.57 to 0.84).

In Sample 2 we also investigated correlations between hostile attributions and the type of scene, vulnerable narcissism was correlated with a greater ascription of intentionally but only in non-intentional scenes (i.e., they perceived hostile intent where none should exist). To demonstrate this result more clearly and examine whether the intentionality attributed in different types of scenes depends on the level of vulnerable narcissism, namely whether vulnerable narcissism predicts the difference between scenes (e.g., Intentionality in Hostile Scenes – Intentionality in Non-Hostile Scenes), we conducted a multiple moderation test with 5,000 bootstrapped samples using MEMORE V2.1. macro (Montoya, 2019). It permits the comparison between two the measurements, hence, we created three models (Hostile – Ambiguous; Hostile – Non-Hostile; Ambiguous – Non-Hostile). In each model, vulnerable narcissism was included as a moderator. There was no moderation for the first model (Hostile – Ambiguous; $F[1, 151] = 0.29$, $p = .59$). We did, however, find moderation for the Hostile – Non-Hostile ($F[1, 151] = 5.65$, $p = .02$) and

[4] In Sample 1, both forms of narcissism were uncorrelated ($r = .07$, $p = .54$), however, there was a significant difference in their severity ($t[73] = 9.44$, $p < .01$), vulnerable narcissism was greater ($M = 3.40$, $SD = 0.60$) than grandiose narcissism ($M = 2.42$, $SD = 0.70$). In Sample 2, both forms of narcissism were also uncorrelated ($r = -.02$, $p = .86$) and there was no differences in their intensity ($t[153] = 1.55$, $p = .12$).

Table 1. Descriptive statistics and correlations between individual differences in hostile attributions and grandiose (NPI) and vulnerable (HSNS) narcissism in Sample 1 and 2

		r	
	M (*SD*)	NPI	HSNS
Sample 1 (*N* = 74)			
Coworker	2.89 (1.11)	−.11	.27*
Authority	3.27 (0.99)	.09	.14
Strangers	2.90 (1.25)	−.23†	.32**
Acquaintance	3.19 (0.98)	.12	.37**
Friend	2.61 (1.15)	.13	.15
Sample 2 (*N* = 153)			
Coworker	2.51 (1.06)	.09	.25**
Authority	3.03 (0.98)	.12	.22**
Strangers	2.11 (1.05)	−.14	.29**
Acquaintance	2.86 (1.01)	.14	.28**
Friend	2.30 (1.19)	.03	.03
Ambiguous	5.60 (0.66)	.08	.02
Hostile	7.40 (0.73)	.06	−.02
Non-Hostile	3.41 (1.05)	−.06	.23**

Note. *M* = Mean; *SD* = Standard Deviation; NPI = Narcissistic Personality Inventory; HSNS = Hypersensitive Narcissism Scale. †$p \leq .10$; *$p \leq .05$; **$p \leq .01$.

Table 2. Vulnerable narcissism (HSNS) as a moderator of the relationships between type of scenes and intentionality

	B (*SE*)	LLCI (95%)	ULCI (95%)
HSNS			
Hostile – Non-Hostile	−0.46 (0.19)*	−0.84	−0.78
Ambiguous – Non-Hostile	−0.41 (0.14)**	−0.68	−0.13
Low (*M* = 2.36)			
Hostile – Non-Hostile	4.24 (0.15)**	3.94	4.53
Ambiguous – Non-Hostile	2.41 (0.11)**	2.20	2.62
Medium (*M* = 2.91)			
Hostile – Non-Hostile	3.99 (0.11)**	3.78	4.20
Ambiguous – Non-Hostile	2.19 (0.08)**	2.04	2.34
High (*M* = 3.45)			
Hostile – Non-Hostile	3.74 (0.15)**	3.44	4.03
Ambiguous – Non-Hostile	1.96 (0.11)**	1.75	2.18

Note. Each level is a ratio of the difference between the intentionality assigned in a particular type of scene. *B* = Unstandardized Beta; *SE* = Standard Error; LLCI = Lower Limit of the Confidence Interval; UPCI = Upper Limit of the Confidence Interval; HSNS = Hypersensitive Narcissism Scale. *$p \leq .05$; **$p \leq .01$.

the Ambiguous – Non-Hostile models (F[1, 151] = 8.66, p = .01) where the former accounted for 3% of the variance and the latter accounted for 5% of the variance. In both cases, vulnerable narcissism moderated the relationship between scene-type and assigned intentionality (Table 2). Post hoc tests indicated that less intentionality was attributed to the Non-Hostile scenes; however, the higher the level of vulnerable narcissism, the smaller the difference between the scenes (Figure 2).

Discussion

In this study, we aimed to investigate whether the relationship between vulnerable narcissism and the tendency to make hostile attributions depended on the socio-relational context of the situation. We also examined if the level of ambiguity of social situations, assessed from the other-referent position, influenced the relationship between vulnerable narcissism and attributing hostile intentionality.

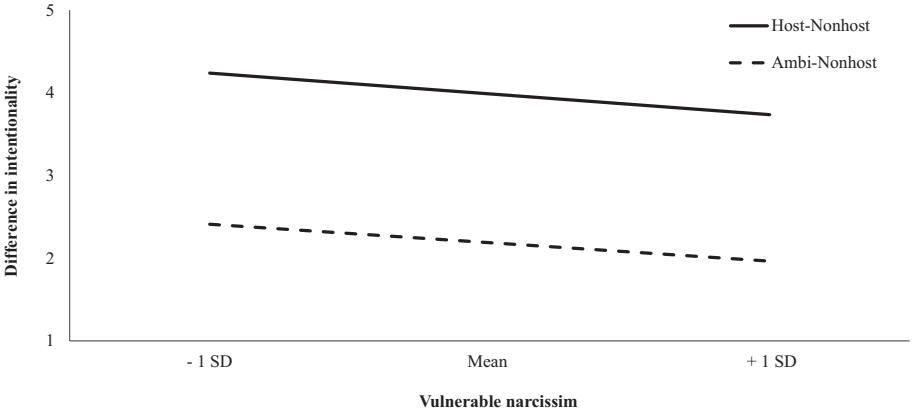


Figure 2. Vulnerable narcissism as a moderator of differences in ascribed intentionality across scene types.

Furthermore, we added some sampling heterogeneity to the field of research on hostile attributions, collecting data from community members engaged in group psychotherapy.

We have largely replicated associations between vulnerable narcissism, but not grandiose narcissism, and hostile attributions, the former of which can be destructive to social interactions (Hansen-Brown & Freis, 2021). Both forms of narcissism lead to interpersonal difficulties, but the sources of these problems are different (Miller et al., 2010). While in grandiose narcissism, disagreeable behaviors towards others result from instrumental motivations (e.g., profit, domination), in the case of vulnerable narcissism, such behaviors result from disagreeableness, stable negative emotionality, and a lack of trust, which increases hostility towards others (Miller et al., 2010, 2018). In general, the current results confirm those obtained earlier, however, they show that these well-described regularities are not fully consistent across situations.

We found that in both clinical and community populations, higher levels of vulnerable narcissism were associated with a greater tendency to infer more hostile intentions in others. Importantly, however, this relationship was absent in the situation with a friend. Friendship is an important kind of relationship for people; perhaps humans naturally look for non-personal, accidental reasons for the behavior of a friend to protect this bond and the self. Vulnerable narcissists could be particularly motivated to do it for fear of rejection. Interestingly, in the clinical group, there was no connection between vulnerable narcissism and hostile attributions bias in the situation with authority. In this item, authority is defined as an "important person," perhaps narcissistic patients have identified with this person, so they have not assessed their actions as hostile. We also verified if there are differences in the strength of the correlation between people from the community and patients who have problems in relationships, which is characteristic of an increased level of vulnerable narcissism, however, we found no differences. The narcissistic personality is dimensional, but in extreme cases, it leads to disordered behaviors (Krizan & Herlache, 2018). In our study, the groups differed in the pattern of the severity of narcissism – the clinical sample had higher levels of vulnerability than grandiose narcissism, while there were no differences in the non-clinical group. Such comparisons are not the purpose of our work here, thus we discuss them with caution. Nevertheless, in subsequent studies, it will be worth verifying whether the imbalance between the two types of narcissism is responsible for interpersonal problems among patients, especially because grandiose narcissism could be conducive to mental health (Kaufman et al., 2020; Sedikides et al., 2004). The need to clarify the within-person dynamics of both types of narcissism was reported in previous works (Wright & Edershile, 2018).

Lastly, we found that vulnerable narcissism was linked to the greater intentionality attributed to actors in non-violent scenes, but not in hostile and ambiguous ones. This result was further expanded by the analysis which showed that individuals with higher levels of narcissism were less likely to differentiate intentionality by scene-type, particularly between Non-hostile scenes (those containing no hostile cue) and Ambiguous or Hostile scenes. It is interesting because an attributional bias primarily occurs in ambiguous social situations (Waas, 1988). That is, in weak situations personality traits take center stage in predicting behaviors. The fact that for people with higher levels of vulnerable narcissism, this bias appears in scenes without hostile cues might indicate a "hypersensitivity" of vulnerable narcissists to perceive threats when others do not see them. Our research, although largely confirming previous results, focuses on the social-relational context (Zajenkowska et al., 2018). Even a vulnerable narcissist may not be sensitive to hostile cues in their relationship with a friend, but when observing social situations from the other-referent position, they may notice hostility even if nothing or no one points to it.

Limitations and Conclusions

Despite the heterogeneity of the samples and the style of assessment of attributions of hostile intent, our study is limited in several ways. First, both samples were problematic: they are somewhat WEIRD (Henrich et al., 2010), imbalanced regarding the participant's sex, they may be contaminated by volunteer bias, and were small given modern standards (Gignac & Szodorai, 2016; Schönbrodt & Perugini, 2013) despite basing our sample sizes on previous research (especially for Sample 1). Second, while we used two different methods of assessing hostile attributions, other methods may be useful. For instance, a recently developed method to assess the perceptions people have of situations was created that relies on situational affordances (Rauthmann & Sherman, 2016). One of those situational perceptions may be relevant here in the form of adversity. However, across two measurement models of narcissism, the researcher did not find a relationship with individual differences in viewing the situations as containing adversity (Jonason & Sherman, 2020; Zajenkowski et al., 2020), however, this work may be capturing grandiose more than vulnerable narcissism. Despite these limitations, we have replicated and extended work suggesting a key difference between vulnerable and grandiose narcissists is that the former see the world and others as more hostile (Hansen-Brown & Freis, 2021).

Of all personality traits under investigation, narcissism has one of the longest research traditions reaching as far back to at least Freud (see Kohut, 1972) if not to the Greek myth from which its name comes from of the man – *Narcissus* – who fell in love with his reflection and then fell to ruin. Traditionally, the trait has been studied solely as a singular pathology, but modern research says there might be two sides of this man (or person): one grandiose and one vulnerable (Dickinson & Pincus, 2003; Sedikides et al., 2004; Wink, 1991). In this study, we add to this modern research tradition, suggesting only one of these faces of *Narcissus* is linked to a tendency to see the others in hostile ways. It was vulnerable, and not grandiose, narcissism that showed any real associations with global hostile attributions of the intentions of others, but these were sensitive to the target in the vignette- or image-stimuli we used. What is called for now, is more research to determine what are the situational and relationship cues – if any – that moderate this hostile bias in vulnerable narcissists.

References

Banaśkiewicz-Bazińska, R., & Drat-Ruszczak, K. (2000). *Struktura narcyzmu w polskiej adaptacji kwestionariusza NPI Ruskina i Halla* [Structure of narcissism in the polish adaptation of the Narcissistic Personality Inventory]. *Czasopismo Psychologiczne, 6*(3–4), 171–188.

Basquill, M. F., Nezu, C. M., Nezu, A. M., & Klein, T. L. (2004). Aggression-related hostility bias and social problem-solving deficits in adult males with mental retardation. *American Journal on Mental Retardation, 109*(3), 255–263. https://doi.org/10.1352/0895-8017(2004)109<255:AHBASP>2.0.CO;2

Combs, D. R., Penn, D. L., Wicher, M., & Waldheter, E. (2007). The Ambiguous Intentions Hostility Questionnaire (AIHQ): A new measure for evaluating hostile social-cognitive biases in paranoia. *Cognitive Neuropsychiatry, 12*(2), 128–143. https://doi.org/10.1080/13546800600787854

Cutrona, C. E., Russell, D., & Jones, R. (1984). Cross-situational consistency in causal attributions: Does attributional style exist? *Journal of Personality and Social Psychology, 47*(5), 1043–1058. https://doi.org/10.1037/0022-3514.47.5.1043

Czarna, A. Z. (2011). Narcyz w sieci społecznej, czyli co analiza sieci mówi o funkcjonowaniu osób narcystycznych [A narcissist in the social network: What does network analysis tell us about functioning of narcissistic individuals?]. *Psychologia Społeczna: Social Psychology, 6*(17), 129–145.

Czarna, A. Z., Dufner, M., & Clifton, A. D. (2014). The effects of vulnerable and grandiose narcissism on liking-based and disliking-based centrality in social networks. *Journal of Research in Personality, 50*, 42–45. https://doi.org/10.1016/j.jrp.2014.02.004

Czarna, A. Z., Zajenkowski, M., Maciantowicz, O., & Szymaniak, K. (2019). The relationship of narcissism with tendency to react with anger and hostility: The roles of neuroticism and emotion regulation ability. *Current Psychology*. Advance online publication. https://doi.org/10.1007/s12144-019-00504-6

Dickinson, K. A., & Pincus, A. L. (2003). Interpersonal analysis of grandiose and vulnerable narcissism. *Journal of Personality Disorders, 17*(3), 188–207. https://doi.org/10.1521/pedi.17.3.188.22146

Eid, M., Gollwitzer, M., & Schmitt, M. (2011). *Statistik und Forschungsmethoden* [Statistics and research methods]. Beltz.

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences, 102*, 74–78. https://doi.org/10.1016/j.paid.2016. 06.069

Hansen-Brown, A. A., & Freis, S. D. (2021). Assuming the worst: Hostile attribution bias in vulnerable narcissists. *Self and Identity, 20*(2), 152–164. https://doi.org/10.1080/15298868.2019.1609574

Hendin, H. M., & Cheek, J. M. (1997). Assessing hypersensitive narcissism: A reexamination of Murray's Narcissism Scale. *Journal of Research in Personality, 31*(4), 588–599. https://doi.org/10.1006/jrpe.1997.2204

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Hepper, E. G., Hart, C. M., Meek, R., Cisek, S., & Sedikides, C. (2014). Narcissism and empathy in young offenders and non-offenders. *European Journal of Personality, 28*(2), 201–210. https://doi.org/10.1002/per.1939

Jahoda, A., Pert, C., & Trower, P. (2006). Frequent aggression and attribution of hostile intent in people with mild to moderate intellectual disabilities: An empirical investigation. *American Journal on Mental Retardation, 111*(2), 90–99. https://doi.org/10.1352/0895-8017(2006)111[90:FAAAOH]2.0.CO;2

Jonason, P. K., & Sherman, R. A. (2020). Personality and the perception of situations: The Big Five and Dark Triad traits. *Personality and Individual Differences, 163*, Article 110081. https://doi.org/10.1016/j.paid.2020.110081

Kaufman, S. B., Weiss, B., Miller, J. D., & Campbell, W. K. (2020). Clinical correlates of vulnerable and grandiose narcissism: A personality perspective. *Journal of Personality Disorders, 34*(1), 107–130. https://doi.org/10.1521/pedi_2018_32_384

Kernberg, O. (1975). *Borderline conditions and pathological narcissism*. Jason Aronson.

Kohut, H. (1972). Thoughts on narcissism and narcissistic rage. *The Psychoanalytic Study of the Child, 27*(1), 360–400. https://doi.org/10.1080/00797308.1972.11822721

Krizan, Z., & Herlache, A. D. (2018). The narcissism spectrum model: A synthetic view of narcissistic personality. *Personality and Social Psychology Review, 22*(1), 3–31. https://doi.org/10.1177/1088868316685018

Krizan, Z., & Johar, O. (2015). Narcissistic rage revisited. *Journal of Personality and Social Psychology, 108*(5), 784–801. https://doi.org/10.1037/pspp0000013

Miller, J. D., Dir, A., Gentile, B., Wilson, L., Pryor, L. R., & Campbell, W. K. (2010). Searching for a vulnerable dark triad: Comparing factor 2 psychopathy, vulnerable narcissism, and borderline personality disorder. *Journal of Personality, 78*(5), 1529–1564. https://doi.org/10.1111/j.1467-6494.2010.00660.x

Miller, J. D., Gentile, B., Wilson, L., & Campbell, W. K. (2013). Grandiose and vulnerable narcissism and the DSM-5 pathological personality trait model. *Journal of Personality Assessment, 95*(3), 284–290. https://doi.org/10.1080/00223891.2012.685907

Miller, J. D., Hoffman, B. J., Gaughan, E. T., Gentile, B., Maples, J., & Campbell, W. K. (2011). Grandiose and vulnerable narcissism: A nomological network analysis. *Journal of Personality, 79*(5), 1013–1042. https://doi.org/10.1111/j.1467-6494.2010.00711.x

Miller, J. D., Lynam, D. R., Vize, C., Crowe, M., Sleep, C., Maples-Keller, J. L., Few, L. R., & Campbell, W. K. (2018). Vulnerable narcissism is (mostly) a disorder of neuroticism. *Journal of Personality, 86*(2), 186–199. https://doi.org/10.1111/jopy.12303

Montoya, A. K. (2019). Moderation analysis in two-instance repeated measures designs: Probing methods and multiple moderator models. *Behavior Research Methods, 51*(1), 61–82. https://doi.org/10.3758/s13428-018-1088-6

Pincus, A. L., & Lukowitsky, M. R. (2010). Pathological narcissism and narcissistic personality disorder. *Annual Review of Clinical Psychology, 6*, 421–446. https://doi.org/10.1146/annurev.clinpsy.121208.131215

Raskin, R., & Hall, C. S. (1979). A narcissistic personality inventory. *Psychological Reports, 45*(2), 590. https://doi.org/10.2466/pr0.1979.45.2.590

Raskin, R., & Terry, H. (1988). A principal-components analysis of the Narcissistic Personality Inventory and further evidence of its construct validity. *Journal of Personality and Social Psychology, 54*(5), 890–902. https://doi.org/10.1037/0022-3514. 54.5.890

Rasmussen, K. (2016). Entitled vengeance: A meta-analysis relating narcissism to provoked aggression. *Aggressive Behavior, 42*(4), 362–379. https://doi.org/10.1002/ab.21632

Rauthmann, J. F., & Sherman, R. A. (2016). Measuring the situational eight DIAMONDS characteristics of situations: An optimization of the RSQ-8 to the S8*. *European Journal of Psychological Assessment, 32*(2), 155–164. https://doi.org/10.1027/1015-5759/a000246

Scheff, T. J. (1987). The shame-rage spiral: A case study of an interminable quarrel. In H. B. Lewis (Ed.), *The role of shame in symptom formation* (pp. 109–149). Erlbaum.

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality, 47*(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009

Sedikides, C., Rudich, E. A., Gregg, A. P., Kumashiro, M., & Rusbult, C. (2004). Are normal narcissists psychologically healthy? Self-esteem matters. *Journal of Personality and Social Psychology, 87*(3), 400–416. https://doi.org/10.1037/0022-3514.87.3.400

Tangney, J. P., Wagner, P. E., Hill-Barlow, D., Marschall, D. E., & Gramzow, R. (1996). Relation of shame and guilt to constructive versus destructive responses to anger across the lifespan. *Journal of Personality and Social Psychology, 70*(4), 797–809. https://doi.org/10.1037/0022-3514.70.4.797

Tosevski, D. L., Milovancevic, M. P., & Gajic, S. D. (2010). Personality and psychopathology of university students. *Current Opinion in Psychiatry, 23*(1), 48–52. https://doi.org/10.1097/YCO.0b013e328333d625

Vize, C. E., Collison, K. L., Crowe, M. L., Campbell, W. K., Miller, J. D., & Lynam, D. R. (2019). Using dominance analysis to decompose narcissism and its relation to aggression and externalizing outcomes. *Assessment, 26*(2), 260–270. https://doi.org/10.1177/1073191116685811

Waas, G. A. (1988). Social attributional biases of peer-rejected and aggressive children. *Child Development, 59*(4), 969–975. https://doi.org/10.2307/1130263

Wilkowski, B. M., Robinson, M. D., Gordon, R. D., & Troop-Gordon, W. (2007). Tracking the evil eye: Trait anger and selective attention within ambiguously hostile scenes. *Journal of Research in Personality, 41*(3), 650–666. https://doi.org/10.1016/j.jrp.2006.07.003

Wink, P. (1991). Two faces of narcissism. *Journal of Personality and Social Psychology, 61*(4), 590–597. https://doi.org/10.1037/0022-3514.61.4.590

Wright, A. G., & Edershile, E. A. (2018). Issues resolved and unresolved in pathological narcissism. *Current Opinion in Psychology, 21*, 74–79. https://doi.org/10.1016/j.copsyc.2017.10.001

Zajenkowska, A., Prusik, M., & Szulawski, M. (2018). What does the ambiguous intentions hostility questionnaire really measure? The importance of context in evaluating hostility bias. *Journal of Personality Assessment, 102*(2), 205–213. https://doi.org/10.1080/00223891.2018.1525389

Zajenkowska, A., & Rajchert, J. (2020). How sensitivity to provocation shapes encoding and interpretation of ambiguous scenes in an eye tracking study. *Journal of Cognitive Psychology, 32*(2), 180–198. https://doi.org/10.1080/20445911.2020.1717498

Zajenkowska, A., Rajchert, J., Macianowicz, O., Holas, P., & Murawic, S. (2019). Cognitive-behavioral (CBT) and psychodynamic (PDT) group psychotherapy and their impact on depressive symptoms and hostile attributions. *International Journal of Group Psychotherapy, 69*(4), 383–407. https://doi.org/10.1080/00207284.2019.1653189

Zajenkowski, M., Jonason, P. K., Leniarska, M., & Kozakiewicz, Z. (2020). Who complies with the restrictions to reduce the spread of COVID-19? Personality and perceptions of the COVID-19 situation. *Personality and Individual Differences, 166*, Article 110199. https://doi.org/10.1016/j.paid.2020.110199

History
Received July 2, 2020
Revision received April 23, 2021
Accepted June 19, 2021
Published online November 10, 2021

Publication Ethics
The study was conducted by the established ethical guidelines and received approval from the authorities of the Wolski Hospital in Warsaw.

Open Data
Data used in this study will be stored at http://www.apsycholab.pl

Funding
The studies were funded by National Science Centre (UMO-2017/26/D/HS6/00258). The second author was partially funded by the Polish National Agency for Academic Exchange (PPN/ULM/2019/1/00019/U/00001).

ORCID
Marta Bodecka-Zych
https://orcid.org/0000-0001-7766-1446

Marta Bodecka-Zych
Psychology Department
The Maria Grzegorzewska University
Szczesliwicka 40
02-353 Warsaw
Poland
mbsd1@aps.edu.pl

Appendix

Ambiguous Intentions and Hostility Questionnaire (AIHQ; Combs et al., 2007)

AIHQ contains five ambiguous situations, as follows:

1. You have been in a new job for 3 weeks. One day, you see one of your new co-workers on the street. You start to walk up to this person and start to say hello, but she/he passes by you without saying hello.
2. You have an appointment with an important person. When you arrive at your appointment, the secretary informs you that the person is not in; they took the day off.
3. You walk past a bunch of teenagers at a mall and you hear them start to laugh.
4. You are supposed to meet a new friend for lunch at a restaurant but she/he never shows up.
5. You call a friend and leave a message on their answering machine, asking them to call you back. One week passes and they have not called you back.

Participants read each scenario, imagined it happening to them, and then used Likert scales to rate whether the other person/s performed the action on purpose (rated from 1 = *definitely no* to 6 = *definitely yes*), how angry it made them feel (rated from 1 = *not at all angry* to 5 = *very angry*), and how much they blamed the other person/s (rated from 1 = *not at all* to 5 = *very much*).

Personality and Individual Differences 192 (2022) 111582

Contents lists available at ScienceDirect

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid

Dad, are they laughing at me? Fathers' vulnerable narcissism and sons' hostile attributions☆

Marta Bodecka-Zych [a,*], Anna Zajenkowska [a], Claire Lawrence [b]

[a] The Maria Grzegorzewska University, Poland
[b] Lawrence PsychAdvisory, United Kingdom of Great Britain and Northern Ireland

ARTICLE INFO

Keywords:
Vulnerable narcissism
Hostile attributions
Fathers
Sons

ABSTRACT

Previous studies have shown that making hostile attributions, that is, attributing hostile intentions to other people mainly in ambiguous situations, increases the likelihood of aggressive behavior. Moreover, fathers' hostile attributions are associated with the aggressiveness of their children. There is also a positive relationship between hostile attributions and vulnerable narcissism among adults. However, currently, there are no empirical studies that test the connections between fathers' VN and the hostile attributions of both fathers and sons. To address this gap, the current study included 77 dyads of sons and fathers. Although fathers' hostile attributions subfactors (intentionality, blame ascription, and angry feelings in ambiguous situations) were not associated with sons' hostile attributions subfactors, fathers' angry feelings, as well as the sons' blame ascription and angry feelings, were related to fathers' vulnerable narcissism. Furthermore, fathers' vulnerable narcissism predicted sons' angry feelings which are, in turn, associated with sons' aggressiveness. The observation that fathers' vulnerable narcissism may have an impact on sons' angry feelings in ambiguous situations, which in turn is related to sons' aggression may be a crucial insight for therapeutic interventions for adolescents with aggression problems and suggests a need to focus on the father-son relationship.

1. Introduction

Hostile attribution (HA) is a "*tendency to interpret the intent of others as hostile when social cues fail to indicate a clear intent*" (Epps and Kendall, 1995, p. 161). HA consists of three subfactors: interpreting the other person's behavior as intentional, blaming others for negative events and accompanying angry feelings (Combs et al., 2007; Zajenkowska et al., 2021). Intentionality and blame occur simultaneously, but are distinct in relation to anger, as intentionality may lead to blame, and blame may lead to anger (Zajenkowska et al., 2021). HA often leads to aggression, specifically reactive (impulsive and retaliatory) aggressive acts (Crick and Dodge, 1994; Dodge, 2006), both in children (especially boys) and in adults (Crain et al., 2005; De Castro et al., 2002; Tuente et al., 2019).

The cycle that leads from social cognition to behavior is described by social information processing model (SIP, Crick and Dodge, 1994), according to which behavior is a function of six steps of processing, including encoding and interpretation of social cues, clarification of goals, response access or construction, response decision and finally behavioral enactment. Each of the steps is based on acquired rules, schemas and social knowledge, which is why previous experiences, including family dynamics, are relevant for processing current events (Crick and Dodge, 1994; Dodge, 2006; Horsley et al., 2010).

1.1. The formation of hostile attributions

The family environment has a significant role in forming the ways in which social information is processed (e.g. Dodge, 2006; Halligan et al., 2007; MacKinnon-Lewis et al., 2001). Ideally, socialization leads to the development of benign attributions, whereby, the intentions of others are typically interpreted as benign or benevolent without evidence to the contrary. However, some children do not acquire such a benign way of interpreting the world and maintain a stable personality trait comprising the tendency to make hostile attributions, that guides behavior and works as a self-fulfilling prophecy (Dodge, 2006). Factors that have been associated with reduced acquisition of benign attributional schemas include insecure attachment relationship, physical abuse, failure in important life tasks and modeling of hostile attributions (Dodge, 2006). Therefore, first caregivers (most often parents) have a

☆ Data used in this study will be stored at: http://www.apsycholab.pl.

* Corresponding author at: The Maria Grzegorzewska University, Institute of Psychology, Szczęśliwicka 40, 02-353 Warsaw, Poland.
E-mail address: mbsd1@aps.edu.pl (M. Bodecka-Zych).

https://doi.org/10.1016/j.paid.2022.111582
Received 1 October 2021; Received in revised form 13 January 2022; Accepted 22 February 2022
Available online 8 March 2022

dual role in supporting the acquisition of a more appropriate or benign attributional default: firstly, they teach through their own interpretations and behaviors, which, according to social learning theory, are imitated by children (MacBrayer et al., 2003); secondly, through attachment, they create a secure or insecure base and schemas for future relationships (Bowlby, 1973). Attachment, understood as an emotional bond between caregiver and infant, provides the basis for working models - internalized beliefs about self and others. The result of working models in which relationships are perceived as insecure and chaotic can be hypersensitivity and unresolved anger, which is related to aggression (Greenberg et al., 1993).

Thus, it appears that children whose parents communicate the attribution of intentionality and blame through their own interpretations of others or are unable to build a sufficiently warm and secure relationship with their child, may have particularly unfavorable conditions for acquiring benign attributional patterns. Furthermore, given the well-described associations of vulnerable narcissism with insecure attachment and hostile attributions bias (Hansen-Brown and Freis, 2021; Miller et al., 2011), developing a benign attributional style can be even more difficult for children of a narcissistic parent.

1.2. Vulnerable narcissism and hostile attributions

Narcissism defined as "*one's capacity to maintain a relatively positive self-image*" is described as both a clinical construct and a personality trait (Pincus and Lukowitsky, 2010). Commonly there are two main subtypes of narcissism: grandiose and vulnerable (e.g. Dickinson and Pincus, 2003). Vulnerable narcissism is associated with neuroticism, shame, hypersensitivity to rejection and criticism (Dickinson and Pincus, 2003) and grandiose narcissism is related more to extraversion, grandiosity and dominance (Miller et al., 2011). Grandiose narcissism is also captured by the diagnostic criteria of narcissistic personality disorder (American Psychiatric Association, 2013). Although, both types share a tendency toward entitled and aggressive interpersonal behavior, only vulnerable narcissism is associated with the hostile attribution of intent (Bodecka-Zych et al., 2021; Hansen-Brown and Freis, 2021). Vulnerable narcissists "*assume the worst*" and perceive the ambiguous behaviors of others as intentional and malevolent (Hansen-Brown and Freis, 2021). Moreover, vulnerable (rather than grandiose) narcissism is associated with fearful attachment style, in which individuals maintain a negative view of themselves and others (Miller et al., 2011).

1.3. Transmission of hostile attributions from parents to children

Although children may learn social perceptions and possibly acquire hostile attributions from their parents (Dodge, 2006), previous research has provided mixed results for this assumption. Halligan et al. (2007) demonstrated that parents' general attributional biases and biases toward their own children are not directly related with children's attributional bias. Their research found that both types of biases are associated with children's externalizing behaviors (Halligan et al., 2007). In turn, another study showed that mothers' hostile attributions are associated with their children's hostile attributions, while fathers' hostile attributions correlate with their children's relational aggression. However, the mechanism by which this relationship works, remains unexplained (Nelson et al., 2008).

In previous studies, researchers usually asked children to evaluate different situations to those presented to their parents (e.g. Halligan et al., 2007; Nelson et al., 2008; MacBrayer et al., 2003). In one study where children and parents were presented with the same vignettes, children were asked to judge the situations from their own perspective, and parents were asked to respond as they thought their child would do. However, in this procedure, the crucial confound is how well parents know their children, or are able to overcome the tendency for projection, which is difficult to measure or control (Nelson et al., 2018). Thus, in the current study, fathers and sons respond on their own behalf, which aims to avoid engaging them in second order theory of mind activities.

1.4. Current study

For years, it was thought that the transmission of HA (conveying a tendency toward hostile attribution) is stronger within gender than across gender (Dodge, 2006; MacBrayer et al., 2003) and the transmission of a tendency toward aggressive behavior may be more likely to occur from fathers to children, rather than from mothers, since males tend to show more overt aggression than females (Moffitt and Caspi, 2001). Indeed, some studies challenge the relevance of the SIP model for girls (Crain et al., 2005). Therefore, in the current study the focus is denoted on the relationship between the father and the son. We examine whether the father's interpretations of situations that could have been faced by their children correspond to the way children interpret the same situations (Questions 1). Previous studies have not considered three subfactors of HA (intentionality, blame ascription and angry feelings, Combs et al., 2007), however, each may play a unique role in the transmission of HA from father to son. Therefore we decided to examine each one separately. Furthermore, authors of previous studies have pointed out that there is a need to look for additional variables associated with the transmission of hostile attributions from father to children (Nelson et al., 2008). As vulnerable narcissism is associated with anger, hostility (Czarna et al., 2019) and hostile attributions (Hansen-Brown and Freis, 2021), and additionally can foster the transfer of HA from narcissistic parent to the child, we tested if the vulnerable narcissism of the father is associated with increased HA in the son (Question 2).

Hostile attributions of parents are associated with their children's externalizing behaviors (Halligan et al., 2007). Considering that anger is important for strengthening the relationship between hostile cognition and aggressive behavior (Crick and Dodge, 1994), we examined also, if fathers' blame ascription and angry feelings, as well as fathers' vulnerable narcissism, predict son's angry feelings in ambiguous harmful situations (Question 3). Perceived intentionality has been shown previously not to be directly related to anger (Zajenkowska et al., 2021), thus it will not be included in the model. Finally, because hostile attributions lead to aggressive behavior, we wanted to explore whether the son's hostile attributions (intentionality, blame ascription, and angry feelings in ambiguous situations) are associated with increased involvement in aggression, both being the perpetrator and also the victim of aggression (Question 4).

2. Material and methods

2.1. Participants and procedure

The current study included 77 dyads of boys ($Range_{Age}$ = 12–18, M_{Age} = 14.53, SD_{Age} = 1.50), and their fathers ($Range_{Age}$ = 34–56, M_{Age} = 44.77, SD_{Age} = 4.28). Most of the fathers held a higher education diploma (62.3%), and were married (97.4%, one person was in an informal relationship, one was divorced). 49.4% respondents' families resided in large cities, 27.3% in towns, 23.4% in villages and had one (14.3%), two (57.1%), three (26.0%), four or five (2.6%) children. Recruitment was conducted online with the invitation to the study posted on social media (via township groups and private profiles). Invitation to participate in the study was also sent to parents at the local schools. In response to the invitation, respondents completed a registration questionnaire in which they provided their e-mail address. In the next step they received a link to a questionnaire containing detailed description of the study and its stages, required consents, contact

information for the investigators and self-report questionnaires (a separate questionnaire was given to the son, a separate one to the father, all the respondents were asked to answer the questions independently) [1]. This was followed by an individual, online meeting with the researcher (held with the son and then with the father separately), which lasted approximately 40 min and was conducted using a program that did not require installation and ensured the anonymity of the subjects. For completing the study, both father and son received a voucher to a popular online store worth $12. The study was approved by the Ethics Committee at the Maria Grzegorzewska University (approval no. 188-2019/2020).

2.2. Measures

To measure fathers' levels of vulnerable narcissism, we used the Polish translation (Czarna et al., 2014) of the Hypersensitive Narcissism Scale (Hendin and Cheek, 1997). Participants were asked to specify their level of agreement (1 = *strongly disagree*; 5 = *strongly agree*) to 10 items (e.g. "*I easily become wrapped up in my own interests and forget the existence of others*" and "*I am secretly 'put out' when other people come to me with their troubles, asking me for my time and sympathy*."). We calculated participants' mean ratings to create a score of vulnerable narcissism. The instrument has acceptable internal consistency ($\alpha = 0.71$).

The sons' aggressiveness was measured with Polish adaptation (Farnicka and Grzegorzewska, 2015) of The Mini Direct and Indirect Aggression Inventory (Mini-DIA; Österman and Björkqvist, 2010). The tool consists of six items measuring on a five-point scale, how often (0 = *never* to 4 = *often*) the participants were the victims and the perpetrators of physical, verbal, and indirect aggression (e.g., "*Recently, how often have you felt like a victim of a physical aggression: another pupil has for instance hit, kicked, or pushed you*"). Three questions regarding being a victim and three questions about being a perpetrator formed two separate scales. For each subscale, we calculated the sum of the responses (Österman and Björkqvist, 2010). The reliability of both scales was acceptable ($\alpha_{victim} = 0.77$; $\alpha_{perpetrator} = 0.69$).

To measure sons' and fathers' subfactors of hostile attributions (intentionality, blame ascription and angry feelings), we asked them to evaluate a series of situations that a teenager may encounter. The first part of this method consists of three vignettes, describing a hypothetical situation in which the causes of the behaviors described are ambiguous in nature. All three scenarios were taken from previous research, where its validity has been demonstrated (e.g. Combs et al., 2007; Crick, 1995; Zajenkowska et al., 2018).

Vignettes were read out loud to participants by the researcher, who asked adolescents to imagine that each situation had happened to them. After each scenario, the young people answered the three questions measuring all HA subfactors – intentionality, blame ascription, and angry feelings. The questions and response scales were taken from The Ambiguous Intentions Hostility Questionnaire (AIHQ, Combs et al., 2007; e.g. [Intentionality] *Do you think the colleague did this on purpose? 1 "definitely no" to 6 "definitely yes"* [Anger] *How angry would this make you? 1 "not at all angry" to 5 "very angry"* [Blame] *How much would you blame the colleague for this? 1 "not at all" to 5 "very much"*). Thereafter, the adolescents were asked to write down two situations, in which someone (e.g. a peer) behaved toward them in a way that caused in them negative feelings. Such a method was also previously used in research of Quigley and Tedeschi (1996, p. 1282); the current study used the same instruction for the participants: "*Think about a specific time someone has harmed you. The harm that occurred could have been physical harm, emotional harm (such as insulting you or betraying you), or economic harm (such as stealing from you). When you have thought of an episode, please answer the following questions concerning that incident*". After describing each situation, boys answered the same three questions assessing intentionality, blame ascription and angry feelings. During the meeting with fathers, the researcher read the same three vignettes as those presented to the sons, and the two incidents written by their sons. Then, the fathers were asked to imagine that those situations had happened to their son. For each vignette, fathers had to answer the same three questions as their sons, but on their own behalf, such as "*Do you as a father:* [Intentionality] *think the colleague did this on purpose,* [Anger] *how angry would that make you, and* [Blame] *how much would you blame the colleague for it*?"

The answers were averaged for all 5 situations, separately to each subfactor, to create scores for intentionality, blame ascription and angry feelings for fathers and sons. Since the task measuring hostile attributions was cross-situational we judge reliability based on average inter-item correlation coefficient (Fathers: $r_{intentionality} = 0.33$; $r_{anger} = 0.48$; $r_{blame} = 0.46$; Sons: $r_{intentionality} = 23$; $r_{anger} = 0.35$; $r_{blame} = 0.29$). The obtained inter-item coefficient values are sufficient, according to commonly used criteria (Clark and Watson, 2016; (Zajenkowska et al., 2021).

2.3. Analytic strategy

All analyses were conducted using IBM SPSS 25.0.0.2 for Windows. The data were encoded in wide format, so that each row represented a father-son pair. Then, we checked the Person's correlations between all study variables, which allowed us to answer the questions: 1, 2 and 4. The power analysis indicates that with the current sample size (77 pairs) and an estimated correlation of 0.3, we should detect results at p = .05 with 85% power. Finally, we conducted regression analysis to answer question 3.

3. Results

Bivariate correlations between all study variables and descriptive statistics are presented in the Table 1. Intentionality, anger and blame attributed by the fathers are positively related with each other, as were intentionality, anger and blame attributed by the sons. Fathers' vulnerable narcissism is positively related to his own anger and to sons' anger and blame. However, fathers' HA subfactors are not directly associated with sons' HA subfactors. Being a victim of aggression is related to being a perpetrator of aggression, and both of these scales are positively related to son's anger.

The results of the hierarchical multiple regression examining fathers' blame ascription, anger and vulnerable narcissism in predicting sons' angry feelings are presented in the Table 2.

Step 1 included fathers' blame [2] and anger scores, in Step 2 we added their vulnerable narcissism scores. The first model proved to be statistically non-significant ($F[2,70] = 1.34$, p = .27). Neither fathers' blame nor anger was significant predictors of sons' anger. However, when fathers' vulnerable narcissism was added in the second step, the model proved to be significant ($F[3,70] = 2.92$, p = .04) and explained 12% of the total sons' anger variance.

4. Discussion

The aim of the current study was to examine the links between fathers' hostile attributions and vulnerable narcissism with their sons' tendency toward HA and aggression. As in previous studies (Halligan et al., 2007; Nelson et al., 2008) we did not find direct relationships between father's and son's attribution patterns. It appears that the modeling process of HA may not sufficiently explain the transmission of

[1] The current study is a part of a greater project, participants completed several personal questionnaires and took part in computer tasks on hostile attributions (the results will be reported elsewhere).

[2] The father's blame does not correlate with the son's angry feelings; however, we include this variable in the model to control the possible effects of blame.

Table 1
Descriptive statistics and correlations between all study variables.

	1	2	3	4	5	6	7	8	9	*M*	*SD*
1. Vulnerable narcissism - Father	–									2.65	0.54
2. Intentionality - Father	−0.00	–								4.06	0.92
3. Anger - Father	0.29*	0.49**	–							2.47	0.82
4. Blame - Father	0.14	0.58**	0.82**	–						2.66	0.87
5. Aggression victim - Son	−0.10	0.07	0.21	0.23	–					2.68	2.07
6. Aggression perpetuator - Son	−0.01	−0.02	0.12	0.13	0.43**	–				2.31	1.76
7. Intentionality - Son	0.21	0.12	0.17	0.10	0.18	−0.09	–			4.22	0.88
8. Anger - Son	0.29*	0.17	0.19	0.13	0.29*	0.30**	0.46**	–		2.87	0.74
9. Blame - Son	0.32**	0.06	0.08	0.03	0.09	0.06	0.60**	75**	–	2.95	0.78

* $p < .05$.
** $p < .01$.

Table 2
Father's blame, anger and vulnerable narcissism in predicting son's angry feelings.

Step	Independent variable	Son's angry feelings		
		β	R^2	ΔR^2
1	Father's blame	−0.08	0.04	0.04
	Father's anger	0.25		
2	Father's blame	0.01	0.12	0.08*
	Father's anger	0.09		
	Father's vulnerable narcissism	0.30*		

* $p < .05$.

HA from father to son. Rather, of greater relevance, are the results concerning the association between the fathers' levels of vulnerable narcissism and sons' HA which may indicate that it is the presence of VN and the attendant parenting practices that foster the sons' higher levels of HA. The results of our study support the thesis that narcissistic father's insecure working model may be internalized by his son and translate into his relationships with peers (e.g. Bowlby, 1973; Mikulincer, 1995). Thus, HA transmission from father to son seems to stem from relationships rather than the simple behavioral modeling.

Father's VN is positively associated with his angry feelings arising in response to situations involving his son; however, there is no significant association with the VN and other HA subfactors - intentionality and blame attributed to the son's peers. Nevertheless, given the positive relationships between VN with angry feelings and the results of previous research (Hansen-Brown and Freis, 2021; Miller et al., 2011), it seems possible that narcissistic fathers assign intentionality and blame, if not toward peers, perhaps toward their own sons. This thesis is consistent with statement: "*To the extent that parents are narcissistic (…) they may demand certain behavior from their children because they see the children as extensions of themselves and need the children to represent them in the world*" (Rappoport, 2005, p. 2). If a son "misrepresented" his father in the social world (e.g. by being potentially disliked or victimized by peers) they may be blamed for this by his narcissistic parent. The son, who is blamed by the narcissistic father, could blame peers and become angry in ambiguous situations, as suggested by the positive correlations between the father's vulnerable narcissism and the son's blame ascription and angry feelings, as well as no direct relationship between HA subfactors of the father and son. In other words, the blame from father to son could be internalized by the son, subsequently externalized to other people, which promotes aggressive behavior (e.g. Quigley and Tedeschi, 1996). However, this assumption needs to be confirmed in further research, in which the father would be asked about the behavioral attributions of both the peer and the son, and the son's negative self-image would be explored as a function of their fathers' VN.

The importance of the Father-Son relationship is also supported by our finding that the fathers' vulnerable narcissism predicts the sons' angry feelings rather than the blame attributed by the father or the fathers' anger. It has been shown that parental narcissism is related with a dysfunctional rearing style and may promote depression and anxiety in children, however this study tested parents' grandiose narcissism (Dentale et al., 2015). There is a gap in the literature regarding the parenting practices of vulnerable narcissists. Nevertheless it appears that high levels of this trait are not conductive to building secure relationships with child (Bowlby, 1973). Overall, narcissistic parents are controlling, blaming, self-absorbed, and do not recognize the needs of their children (Rappoport, 2005). Moreover, previous research has demonstrated that there is a link between insecure attachment and various aspects of anger (e.g. Brodie et al., 2019). Therefore, we suspect that the relationship between the father's vulnerable narcissism and the son's angry feelings is mediated by the quality of the attachment relationship, which requires confirmation in future research.

Finally, as we expected, the sons' angry feelings in response to an ambiguous situation were associated with being both a victim and a perpetrator of aggression. Precursors of SIP research speculated that the links between HA and aggression are the strongest for children who react with anger as a response to a potential threat (Crick and Dodge, 1994; Dodge, 2006). However, it is important to remember that the hostile attributions could be both a cause and a consequence of peer troubles (e.g. Dodge and Pettit, 2003), and that peers may also be a source of a young person developing HA (Dodge, 2006). Therefore, on the one hand, anger can lead to problems with peers; on the other hand, negative experiences with peers can lead to anger in future situations, hence these results should be treated with caution.

5. Limitations and conclusions

Our findings should be interpreted in the light of some limitations. First, the nature of the data collected (relatively small sample size, VN measure for fathers only, aggression measure for sons only), did not allow us to conduct an analysis that takes into account the dyadic nature of the data (e.g. APIM; Cook and Kenny, 2005). However, we would like to emphasize that the study of bidirectional effects in interpersonal relationships was not our primary goal. A second limitation refers to the single-item measurements of anger, blame and attributed intentionality in each scene which simultaneously were the subfactors of hostile attributions. Diverse methods of measuring HA, and factors strictly related to this tendency could be included in future studies, as the additional items may enable the capture of more variance of the variable. In future research, it might also useful to control for the actual quality of the father-son relationship, e.g., positive time spent together, which may be crucial for both modeling behavior and forming a secure relationship.

Despite some limitations, the current study expands the body of knowledge behind hostile attributions, primarily suggesting that the cognitive pattern model may not be the primary mechanism for the transmission of hostile attributions from father to son. It seems that sons acquire a hostile attributional style less by observing narcissistic fathers, and more by being in a relationship with them, and therefore through developing a self-image and self-esteem in this relationship and generating hostility toward the potential source of threat. This may have

practical implications for therapeutic work with adolescents with aggression problems.

Funding

This work was supported by Ministry of Science and Higher Education in Poland in the form of subsidy for the maintenance and development of research potential at The Maria Grzegorzewska University (grant no. BNS 54/21-P) and National Science Centre of Poland (grant no. UMO-2017/26/D/HS6/00258).

CRediT authorship contribution statement

Marta Bodecka-Zych: Conceptualization, Methodology, Formal analysis, Investigation, Writing – Original draft, Writing – Review & ending, Project administration, Funding acquisition.
Anna Zajenkowska: Conceptualization, Methodology, Formal analysis, Writing – Original draft, Writing – Review & ending, Supervision, Funding acquisition.
Claire Lawrence: Writing – Original draft, Writing – Review & ending, Supervision.

References

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.).

Bodecka-Zych, M., Jonason, P. K., & Zajenkowska, A. (2021). Hostile attribution biases in vulnerable narcissists depends on the socio-relational context. *Journal of Individual Differences*. https://doi.org/10.1027/1614-0001/a000354. Advance online publication.

Bowlby, J. (1973). Attachment and loss: Volume II: Separation, anxiety and anger. In *Attachment and loss: Volume II: Separation, anxiety and anger* (pp. 1–429). London: The Hogarth press and the institute of psycho-analysis.

Brodie, Z. P., Goodall, K., Darling, S., & McVittie, C. (2019). Attachment insecurity and dispositional aggression: The mediating role of maladaptive anger regulation. *Journal of Social and Personal Relationships, 36*(6), 1831–1852.

Clark, L. A., & Watson, D. (2016). Constructing validity: Basic issues in objective scale development. In A. E. Kazdin (Ed.), *Methodological issues and strategies in clinical research* (pp. 187–203). Worcester, MA: American Psychological Association.

Combs, D. R., Penn, D. L., Wicher, M., & Waldheter, E. (2007). The ambiguous intentions hostility questionnaire (AIHQ): A new measure for evaluating hostile social-cognitive biases in paranoia. *Cognitive Neuropsychiatry, 12*(2), 128–143.

Cook, W. L., & Kenny, D. A. (2005). The actor–partner interdependence model: A model of bidirectional effects in developmental studies. *International Journal of Behavioral Development, 29*(2), 101–109.

Crain, M. M., Finch, C. L., & Foster, S. L. (2005). The relevance of the social information processing model for understanding relational aggression in girls. *Merrill-Palmer Quarterly, 51*(2), 213–249.

Crick, N. R. (1995). Relational aggression: The role of intent attributions, feelings of distress, and provocation type. *Development and Psychopathology, 7*(2), 313–322.

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin, 115*(1), 74–101.

Czarna, A. Z., Dufner, M., & Clifton, A. D. (2014). The effects of vulnerable and grandiose narcissism on liking-based and disliking-based centrality in social networks. *Journal of Research in Personality, 50*, 42–45.

Czarna, A. Z., Zajenkowski, M., Maciantowicz, O., & Szymaniak, K. (2019). The relationship of narcissism with tendency to react with anger and hostility: The roles of neuroticism and emotion regulation ability. *Current Psychology*, 1–16.

De Castro, B. O., Veerman, J. W., Koops, W., Bosch, J. D., & Monshouwer, H. J. (2002). Hostile attribution of intent and aggressive behavior: A meta-analysis. *Child Development, 73*(3), 916–934.

Dentale, F., Verrastro, V., Petruccelli, I., Diotaiuti, P., Petruccelli, F., Cappelli, L., & San Martini, P. (2015). Relationship between parental narcissism and children's mental vulnerability: Mediation role of rearing style. *International Journal of Psychology and Psychological Therapy, 15*(3), 337–347.

Dickinson, K. A., & Pincus, A. L. (2003). Interpersonal analysis of grandiose and vulnerable narcissism. *Journal of Personality Disorders, 17*(3), 188–207.

Dodge, K. A. (2006). Translational science in action: Hostile attributional style and the development of aggressive behavior problems. *Development and Psychopathology, 18*(3), 791–814.

Dodge, K. A., & Pettit, G. S. (2003). A biopsychosocial model of the development of chronic conduct problems in adolescence. *Developmental Psychology, 39*(2), 349.

Epps, J., & Kendall, P. C. (1995). Hostile attributional bias in adults. *Cognitive Therapy and Research, 19*(2), 159–178.

Farnicka, M., & Grzegorzewska, I. (2015). Intrapersonal correlates of aggression in adolescents: Determinants of undertaking the role of the perpetrator and the victim. *Current Issues in Personality Psychology, 3*(1), 25–35.

Greenberg, M. T., Speltz, M. L., & Deklyen, M. (1993). The role of attachment in the early development of disruptive behavior problems. *Development and Psychopathology, 5*(1–2), 191–213.

Halligan, S. L., Cooper, P. J., Healy, S. J., & Murray, L. (2007). The attribution of hostile intent in mothers, fathers and their children. *Journal of Abnormal Child Psychology, 35*(4), 594–604.

Hansen-Brown, A. A., & Freis, S. D. (2021). Assuming the worst: Hostile attribution bias in vulnerable narcissists. *Self and Identity, 20*(2), 152–164.

Hendin, H. M., & Cheek, J. M. (1997). Assessing hypersensitive narcissism: A reexamination of Murray's narcism scale. *Journal of Research in Personality, 31*(4), 588–599.

Horsley, T. A., de Castro, B. O., & Van der Schoot, M. (2010). In the eye of the beholder: Eye-tracking assessment of social information processing in aggressive behavior. *Journal of Abnormal Child Psychology, 38*(5), 587–599.

MacBrayer, E. K., Milich, R., & Hundley, M. (2003). Attributional biases in aggressive children and their mothers. *Journal of Abnormal Psychology, 112*(4), 698.

MacKinnon-Lewis, C., Castellino, D. R., Brody, G. H., & Fincham, F. D. (2001). A longitudinal examination of the associations between fathers' and children's attributions and negative interactions. *Social Development, 10*(4), 473–487.

Mikulincer, M. (1995). Attachment style and the mental representation of the self. *Journal of Personality and Social Psychology, 69*(6), 1203–1215.

Miller, J. D., Hoffman, B. J., Gaughan, E. T., Gentile, B., Maples, J., & Keith Campbell, W. (2011). Grandiose and vulnerable narcissism: A nomological network analysis. *Journal of Personality, 79*(5), 1013–1042.

Moffitt, T. E., & Caspi, A. (2001). Childhood predictors differentiate life-course persistent and adolescence-limited antisocial pathways among males and females. *Development and Psychopathology, 13*(2), 355–375.

Nelson, D. A., Cramer, C. M., Coyne, S. M., & Olsen, J. A. (2018). Children's hostile intent attributions and emotional distress: What do parents perceive? *Aggressive Behavior, 44*(1), 98–108.

Nelson, D. A., Mitchell, C., & Yang, C. (2008). Intent attributions and aggression: A study of children and their parents. *Journal of Abnormal Child Psychology, 36*(6), 793–806.

Österman, K., & Björkqvist, K. (2010). The mini direct indirect aggression inventory (Mini-DIA). In K. Österman (Ed.), *Indirect and direct aggression* (pp. 103–111). Frankfurt am Main: Peter Lang, Germany.

Pincus, A. L., & Lukowitsky, M. R. (2010). Pathological narcissism and narcissistic personality disorder. *Annual Review of Clinical Psychology, 6*(1), 421–446.

Quigley, B. M., & Tedeschi, J. T. (1996). Mediating effects of blame attributions on feelings of anger. *Personality and Social Psychology Bulletin, 22*(12), 1280–1288.

Rappoport, A. (2005). Co-narcissism: How we accommodate to narcissistic parents. *The Therapist, 1*, 1–8.

Tuente, S. K., Bogaerts, S., & Veling, W. (2019). Hostile attribution bias and aggression in adults-a systematic review. *Aggression and Violent Behavior, 46*, 66–81.

Zajenkowska, A., Prusik, M., & Szulawski, M. (2018). What does the ambiguous intentions hostility questionnaire really measure? The importance of context in evaluating hostility bias. *Journal of Personality Assessment, 102*(2), 205–213.

Zajenkowska, A., Prusik, M., Jasielska, D., & Szulawski, M. (2021). Hostile attribution bias among offenders and non-offenders: Making social information processing more adequate. *Journal of Community & Applied Social Psychology, 31*(2), 241–256.

PSYCHOLOGIA WYCHOWAWCZA
NR 20/2021, 101–115

ISSN 0033-2860
DOI: 10.5604/01.3001.0015.3825

Data wpływu: 06.2021
Data przyjęcia: 06.2021

MARTA BODECKA-ZYCH[1], ADRIANNA JAKUBOWSKA[1],
ANNA ZAJENKOWSKA[2]

[1] Akademia Pedagogiki Specjalnej im. Marii Grzegorzewskiej, Szkoła Doktorska,
[2] Akademia Pedagogiki Specjalnej im. Marii Grzegorzewskiej, Instytut Psychologii

WARSZTAT PSYCHOEDUKACYJNY Z ELEMENTAMI MENTALIZACJI I JEGO ROLA W REDUKOWANIU WROGOŚCI WŚRÓD OSÓB OSADZONYCH

Streszczenie: Według teorii przetwarzania informacji społecznych zachowanie jest wypadkową sześciu następujących po sobie kroków. Pierwsze dwa – kodowanie i interpretacja – są kluczowe, ponieważ mogą decydować o tym, jak przebiegną kolejne etapy procesu. Adekwatne interpretowanie zachowań własnych i innych ludzi jest możliwe dzięki rozwiniętej zdolności do mentalizowania. Mentalizacja jest wyobrażeniową aktywnością umysłową, która ma na celu uświadomienie ludzkiego zachowania jako intencjonalnych stanów umysłu. Niepełna lub nieprawidłowa mentalizacja utrudnia efektywne funkcjonowanie w świecie społecznym, może np. zwiększać ryzyko wystąpienia zachowań agresywnych. Dotychczasowe badania wykazały, że osoby osadzone wykazują zmniejszoną zdolność do mentalizowania. Jednocześnie jest to umiejętność, która może być rozwijana na każdym etapie życia. W związku z tym w artykule przedstawiamy także warsztat psychoedukacyjny z elementami mentalizacji. Opisujemy jego przebieg, ale także dotychczasowe doświadczenia osób prowadzących (liderów) oraz wstępne wyniki badań potwierdzających skuteczność metody.

Słowa kluczowe: mentalizacja, przywiązanie, osadzeni, przetwarzanie informacji społecznych

Przetwarzanie informacji społecznych

Wyobraź sobie taką sytuację: Idziesz ulicą, mijasz grupę nieznanych Ci osób, nagle słyszysz, że zaczynają się śmiać. Co przychodzi Ci do głowy? Jaka byłaby Twoja pierwsza myśl? Czy te osoby śmieją się z Ciebie? A może jedna z nich opowiedziała żart?

Sposób, w jaki postrzegamy i interpretujemy sytuacje społeczne, wiąże się z naszym zachowaniem. Według społeczno-poznawczego modelu przetwarzania informacji społecznych (*Social Information Processing*, SIP) autorstwa Cricka i Dodge'a (1994) zachowania społeczne są konsekwencją sześciu następujących po sobie kroków (rysunek 1). Przetwarzanie rozpoczyna się od kodowania dostępnych informacji, które mogą pochodzić

Adres do korespondencji: Marta Bodecka-Zych, mbsd1@aps.edu.pl, ORCID: https://orcid.org/0000-0001-7766-1446; Adrianna Jakubowska, aj62194@aps.edu.pl, ORCID: https://orcid.org/0000-0002-8778-0067; Anna Zajenkowska, azajenkowska@aps.edu.pl, ORCID: https://orcid.org/0000-0002-9413-1428.

z różnych kanałów sensorycznych i mieć zarówno werbalny, jak i niewerbalny charakter (Beer, Ochsner, 2006). Zakodowane informacje (czyli te, które dotarły do odbiorcy) stanowią dane wyjściowe do dalszego procesu przetwarzania i wskazówki dla interpretacji, która jest drugim etapem w modelu. Ta sama wskazówka może zostać zinterpretowana na różne sposoby, np. zmarszczone brwi mogą zostać odebrane jako oznaka gniewu lub koncentracji (Beer, Ochsner, 2006). Na tym etapie są podejmowane też wstępne decyzje dotyczące intencjonalności działań innych osób (Dodge, 1986). W interpretowaniu poszczególnych wskazówek ważną rolę odgrywa szeroko rozumiany kontekst sytuacyjny, w którym znajduje się dana osoba (zwana też aktorem), a także jej wcześniejsze doświadczenia. W trzecim kroku aktor określa swój cel lub pożądany wynik sytuacji. Takim celem może być uniknięcie kłopotów, zdobycie przyjaciela, a także okazanie dominacji, zachowanie twarzy czy obrona (Crick, Dodge, 1994). Kolejnym krokiem jest określenie możliwych reakcji w odpowiedzi na daną sytuację. W tym celu aktorzy poszukują mentalnie różnych możliwości w pamięci długotrwałej, a w przypadku nowej sytuacji konstruują odpowiedź i wybierają tę najbardziej według nich odpowiednią. Ostatni krok to realizacja podjętej decyzji w formie konkretnego zachowania.

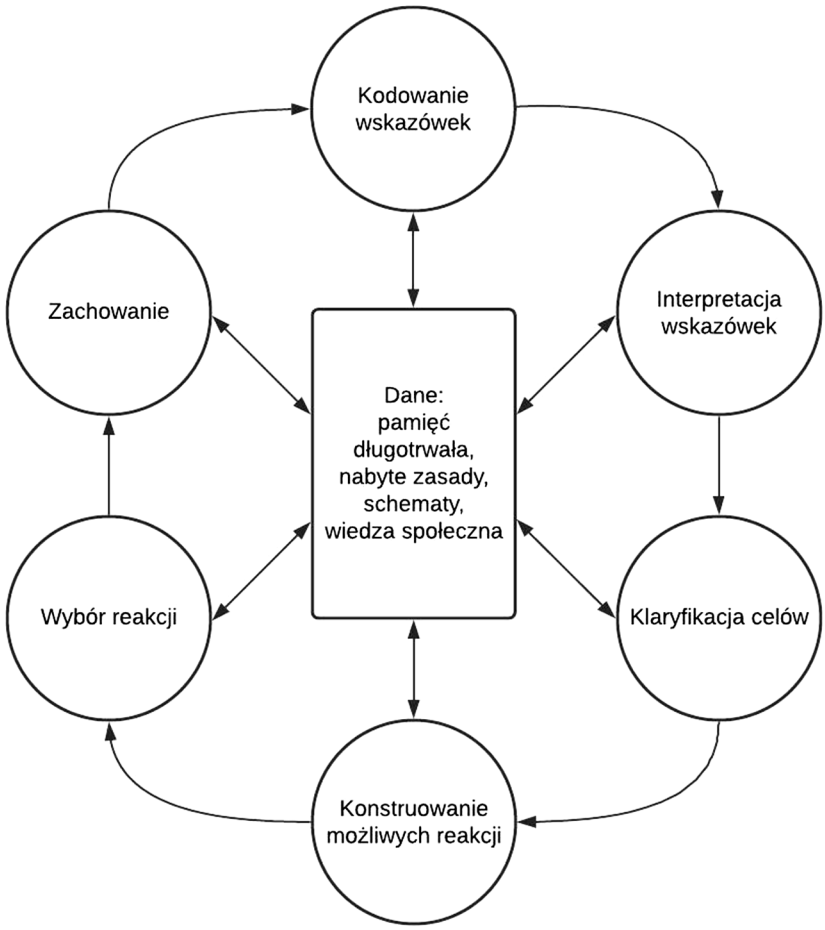
Co istotne, poszczególne etapy przetwarzania informacji społecznych następują niemalże równolegle, najczęściej mają charakter automatyczny (przebiegają bez udziału świadomości) i podlegają wpływowi emocji (Crick, Dodge, 1994).

Adekwatne przetwarzanie informacji na każdym z tych etapów pozwala na efektywne funkcjonowanie w sytuacjach społecznych, podczas gdy niepełne lub tendencyjne przetwarzanie może prowadzić do niepożądanych zachowań – w tym do zachowań agresywnych (Crick, Dodge, 1996). Pierwotnie teoria przetwarzania informacji społecznych odnosiła się głównie do dzieci i młodzieży (a właściwie nastoletnich chłopców). Crick i Dodge (1994, 1996) postawili hipotezę, że agresywność wśród nastolatków wiąże się ze sposobem, w jaki odbierają oni rzeczywistość. Ponadto zakładali, że agresywni chłopcy zwracają uwagę na dostępne informacje w sposób selektywny, np. poświęcają więcej uwagi wrogim wskazówkom, w związku z tym często interpretują sytuacje społeczne, korzystając w dużej mierze z wcześniej nabytych schematów (Gerrig, 1988).

Dzieci, w szczególności chłopcy (Crain i in., 2005) przejawiający zachowania agresywne, dokonują wrogich atrybucji w odpowiedzi na niejednoznaczne sytuacje społeczne, to znaczy interpretują zachowania innych ludzi jako intencjonalnie krzywdzące (Dodge, 2006). W wyniku tego typu interpretacji może pojawić się agresja reaktywna, będąca odpowiedzią na postrzegane zagrożenie. Związek pomiędzy przypisywaną wrogością a gniewem i agresją w sytuacjach ambiwalentnych został potwierdzony także wśród osób dorosłych, jednak autorzy zwracają uwagę na uwzględnienie różnic międzypłciowych (Epps, Kendall, 1995; Zajenkowska, Rajchert, 2020).

Kolejne badania dostarczyły dowodów afirmujących podstawowe założenia teorii przetwarzania informacji społecznych, chociaż zostały one w dużej mierze uszczegółowione. Np. badania z wykorzystaniem technologii eyetrackingu (śledzenia ruchów gałek ocznych) wykazały, że podczas kodowania informacji dzieci agresywne nie poświęcają więcej uwagi wrogim wskazówkom – wręcz przeciwnie – dłużej przyglądają się wskazówkom świadczącym o nieintencjonalnym działaniu, a mimo to interpretują zachowanie aktora jako wrogie i zamierzone. Wyniki te świadczą o odgórnym przetwarzaniu informacji społecznych, w którym ważną rolę odgrywają posiadane schematy. Agresywne dzieci już na etapie kodowania polegają na wcześniej nabytych strukturach wiedzy – poświęcają więcej uwagi wskazówkom niepasującym do schematu, mimo tego przypisują wrogie intencje innym osobom (Horsley i in., 2010).

Rysunek 1
Model przetwarzania informacji społecznych (Crick, Dodge, 1994, s. 76)



W związku z tym w procesie przetwarzania informacji o świecie społecznym ważną rolę odgrywają wcześniej nabyta wiedza i doświadczenia, a te zaczynamy zdobywać bardzo wcześnie – już podczas pierwszej relacji z opiekunem w okresie niemowlęctwa (Bowlby, 2007). Podczas tych wczesnodziecięcych relacji kształtują się wewnętrzne modele robocze – oparte na doświadczeniu reprezentacje mentalne, które pomagają w gromadzeniu i interpretowaniu informacji związanych z szeregiem czynników społecznych (Bowlby, 1973). Modele te, w zależności od jakości relacji, mogą stanowić bezpieczny lub pozabezpieczny wzór przywiązania, a tym samym prowadzić do adaptacyjnego bądź nieadaptacyjnego przetwarzania informacji społecznych (Dykas, Cassidy, 2011).

Osoby charakteryzujące się bezpiecznym wzorcem przywiązania przetwarzają informacje społeczne w sposób pełny i elastyczny, biorą pod uwagę szeroki zakres wskazówek, zarówno pozytywnych, jak i negatywnych, korzystają ponadto ze swoich doświadczeń, aby tworzyć łagodne atrybucje wobec innych ludzi; z kolei osoby charakteryzujące się pozabezpiecznym wzorcem mogą przetwarzać informacje społeczne

na jeden z dwóch sposobów: defensywnie unikać informacji, które mogą je zranić, lub przetwarzać te informacje w sposób wrogi – zgodny z negatywnym schematem (Dykas, Cassidy, 2011). Tym, co odpowiada za różnice w przetwarzaniu informacji społecznych przez osoby z bezpiecznym i pozabezpiecznym stylem przywiązania, może być zróżnicowana zdolność do mentalizowania, czyli odczytywania stanów umysłowych – własnych i innych ludzi (Fonagy i in., 2011).

Czym jest mentalizacja?

Mentalizacja to konstrukt wieloaspektowy, modyfikowany i rozbudowywany w ciągu lat (Adamczyk, 2013a, 2013b), występujący także pod nazwą refleksyjności (ang. *reflective functioning*; Fonagy, Levinson, 2004) oraz zestawiany z pojęciami takimi jak uważność (ang. *mindfulness*), empatia (Choi-Kain, Gunderson, 2008) czy metapoznanie (ang. *metacognition*; Papaleontiou-Louca, 2008). Pojęcie to definiowane jest jako zdolność lub aktywność umysłowa pozwalająca jednostce brać pod uwagę stan, pragnienia i cele osoby, której zachowanie poddaje interpretacji (Coates, 2006, s. 16). Osoba, która mentalizuje, jest świadoma, że zachowanie człowieka (swoje i innych ludzi) wynika z jego stanu mentalnego (Bateman, Fonagy, 2004). Mentalizacja wiąże się zarówno z elastycznym i nieustannym interpretowaniem sygnałów emocjonalnych płynących od innych, jak i z regulowaniem emocji własnych (Bateman, Fonagy, 2004). Pozwala ona trafniej odczytywać intencje drugiego człowieka, dzięki czemu wpływa na lepsze rozumienie sytuacji społecznych i na wzmocnienie relacji interpersonalnych przez interpretującego (Adamczyk, 2013a, 2013b). Zdolność ta wpływa więc na lepsze funkcjonowanie jednostki w społeczeństwie (Luyten i in., 2020).

Ludzie zwykle mentalizują w sposób automatyczny, przy pewnym wysiłku mogą jednak „przełączyć się" na tryb refleksyjny, w którym świadomie myślą o emocjach, motywacjach czy zachowaniach własnych i innych ludzi (wymiar mentalizacji automatycznej-kontrolowanej). Dodatkowo mogą mentalizować na temat własnych zachowań i działań, ale także myśleć o tym, dlaczego inni ludzie zachowują się w jakiś określony sposób (wymiar self–inni). Ponadto można się zastanawiać, jakie są motywy czyjegoś zachowania, ale także jakie emocje skłoniły mnie lub kogoś do konkretnego działania (wymiar poznawczy-afektywny). Ciekawe w przypadku mentalizacji jest także to, skąd czerpiemy wiedzę na temat stanów mentalnych, czy myślimy, że ktoś nas nie lubi, bo „krzywo na nas patrzy", czy też może dlatego, że nie chce się z nami spotkać (wymiar skupienia na cechach wewnętrznych – skupienia na cechach zewnętrznych) (Luyten i in., 2012).

Współczesne rozumienie mentalizacji nawiązuje do teorii Fonagy'ego (1991) opartej o teorię przywiązania, która bazuje na założeniu, że kształtowanie się poczucia siebie (ang. *sense of self*) dziecka zależy od odpowiedzi opiekunów na jego zachowanie (Bateman, Fonagy, 2004). Jeśli opiekunowie adekwatnie reagują na sygnały dziecka, to jest wyraźnie odzwierciedlają jego stany psychiczne, dziecko ma możliwość zrozumienia, że odzwierciedlane emocje są jego – nie opiekuna. Dziecko uczy się, że świat przez niego postrzegany jest jego interpretacją, a nie odzwierciedleniem rzeczywistości. W efekcie rozwija się u niego poczucie siebie i rozumienie odrębności od innych ludzi (Adamczyk, 2013a, 2013b). Można wynieść z tego wniosek, że właściwe reakcje opiekunów na zachowanie dziecka przyczyniają się do ukształtowania u niego prawidłowej zdolności mentalizowania.

W teorii mentalizacji uważa się, że reakcje opiekunów na zachowanie dziecka wraz z jego umiejętnościami wrodzonymi pełnią główną rolę w początkowym kształtowaniu się mentalizacji (Fonagy i in., 2011). Tym samym można stwierdzić, że rodzicielstwo pozbawione wspomnianych reakcji lub w nie ubogie zagraża rozwojowi poczucia siebie i rozwojowi mentalizacji dziecka (Fonagy i in., 2004). Nierozwinięta zdolność do mentalizacji zniekształca jednostce jej obraz siebie i innych (Bateman, Fonagy, 2004), co przekłada się na zwiększenie niespójności wyrażanych przez nią emocji (Fonagy i in., 2011). Nierozwinięta zdolność do mentalizacji negatywnie wpływa na zdrowie człowieka (np. Bateman i Fonagy, 2013; Fonagy i in., 2011), a deficyty w jej zakresie uważane są za jedne z możliwych sygnałów zaburzeń emocjonalnych, zaburzeń odżywiania oraz zaburzeń osobowości, takich jak np. zaburzenie osobowości typu borderline i antyspołeczne zaburzenie osobowości (Adamczyk, 2013a, 2013b).

W przypadku zaburzeń osobowości deficyty mentalizacyjne uwidaczniają się w zależności od kontekstu, stąd mentalizacja jest też nazywana dynamicznie rozwijającym się stanem (Cierpiałkowska, Górska, 2016). Sami badacze mentalizacji przyjęli założenie, że zdolność do mentalizacji podlega ciągłym wpływom środowiska, w tym jest związana z aktualnie doświadczanymi przez jednostkę emocjami (Fonagy i in., 2011). Wysuwa się wobec tego wniosek, że zdolność do mentalizacji zależy od sytuacji, w jakiej znajduje się jednostka. Środowisko może mieć zarówno negatywny, jak i pozytywny wpływ na rozwój tejże zdolności na każdym etapie życia człowieka. Potwierdzają to wyniki terapii metodą powstałą na gruncie teorii Fonagy'ego (1991), to jest metodą opartą na mentalizacji (ang. *mentalization-based treatment*, MBT), które wykazały jej skuteczność w przypadku leczenia pacjentów, m.in. z objawami depresyjnymi i lękowymi. Badania potwierdziły również jej pozytywny wpływ na społeczne i interpersonalne funkcjonowanie pacjentów. Co ważne, wpływ ten zwiększał się w ciągu 18-miesięcznej obserwacji (Bateman, Fonagy, 2001). Mentalizacja jest więc zdolnością, która w wyniku odpowiednich i korektywnych działań może rozwinąć się na każdym etapie życia człowieka (Zajenkowska i in., 2019, 2021).

Rola mentalizacji wśród osób osadzonych skazanych za przestępstwa z użyciem przemocy

Według niektórych teorii psychologicznych agresywne impulsy są częścią ludzkiej natury, pojawiają się we wczesnym dzieciństwie i umożliwiają biologiczne przetrwanie (Dodge, 1980, 2006; Fonagy, 2003; Hay, 2005). W toku socjalizacji i dzięki budowaniu więzi z innymi ludźmi uczymy się je kontrolować. Czasem, w wyniku splotu różnych czynników – zarówno biologicznych (np. Caspi, 2000), jak i środowiskowych (np. Farrington i in., 2001) – agresywny wzorzec zachowań jest utrwalany, co może wiązać się z wieloma negatywnymi konsekwencjami – problemami w relacjach z innymi ludźmi, trudnością w regulacji emocji, a w skrajnych przypadkach z wkroczeniem na ścieżkę przestępczą.

Więźniowie skazani za przestępstwa z użyciem przemocy postrzegani są jako osoby charakteryzujące się podwyższonym poziomem gniewu i agresywności (np. Suter i in., 2002). Oznacza to, że wykazują oni ogólną dyspozycję do intensywnego i częstego odczuwania gniewu oraz reagowania w sposób agresywny (Spielberger i in., 1988). Emocje, takie jak wspomniany gniew, odgrywają rolę pośredniczącą pomiędzy atrybucją a zachowaniem – zinterpretowanie zamiarów drugiej osoby jako wro-

gich może w zasadzie wywołać dwa rodzaje emocji – strach lub gniew (Crick, Dodge, 1994). Podczas gdy strach prowadzi raczej do społecznego wycofania, reaktywny gniew sprzyja zachowaniom agresywnym (Dodge, 2006). Osoby osadzone, skazane za przestępstwa z użyciem przemocy wykazują szczególne trudności w regulacji emocji (Roberton i in., 2014). Co więcej, badania prowadzone przez Fonagy'ego i Levinson (2004) wykazały, że więźniowie (w porównaniu z osobami z diagnozą zaburzeń osobowości i osobami z grupy kontrolnej) posiadają zmniejszoną zdolność do mentalizowania na temat własnych lub cudzych stanów psychicznych. Zarówno trudności w regulacji emocji, jak i kłopoty związane z mentalizacją wynikają często z trudnych, wczesnodziecięcych doświadczeń.

Fonagy i Levinson (2004) uważają, że wczesnodziecięca trauma związana z przywiązaniem prowadzi do rozwojowej psychopatologii, polegającej na wypieraniu tych doświadczeń i unikaniu myślenia o nich, co prowadzi do deficytów w zdolności do mentalizowania – osoba doświadcza stanów psychicznych w sposób fizyczny, jest nastawiona na obronę, szczególnie poprzez akty przemocy. Obszerna metaanaliza przeprowadzona w 2014 r. przez Ogilvie i współpracowników potwierdziła, że pozabezpieczne przywiązanie jest związane ze wszystkimi rodzajami czynów karalnych, szczególnie silnie z tymi z użyciem przemocy. Deficyty w mentalizowaniu mogą więc pośredniczyć w relacji między pozabezpiecznym przywiązaniem a przestępczością (Fonagy i Levinson, 2004; Ogilvie i in., 2014). Wydaje się, że rozwijanie zdolności mentalizacji wśród osób osadzonych może skutecznie przyczynić się do poprawy ich funkcjonowania w świecie społecznym, a tym samym być skutecznym narzędziem służącym resocjalizacji.

Warsztat psychoedukacyjny oparty na mentalizacji

Warsztat służący poprawie zdolności mentalizacyjnych opiera się na programie Mentalization-Based Treatment (MBT), który został stworzony przez Batemana i Fonagy'ego (2013)[1]. Program ten był początkowo stosowany głównie w pracy z pacjentami z zaburzeniem osobowości typu borderline, (BPD). Obecnie jest jednak wykorzystywany wśród wielu różnych grup: wśród osób z antyspołecznym zaburzeniem osobowości, zaburzeniami odżywiania, uzależnionymi od substancji psychoaktywnych, w pracy z nastolatkami i ich rodzinami (Asen, Fonagy, 2011; Bateman, Fonagy, 2011).

Oryginalny program terapeutyczny przeznaczony dla pacjentów z zaburzeniami osobowości składa się z psychoterapii indywidualnej i grupowej trwającej półtora roku. Zanim jednak pacjenci do niej dołączą, biorą udział w warsztacie psychoedukacyjnym, który składa się z 12 sesji i może być stosowany jako samodzielny blok (Volkert i in., 2019). Warsztat odnosi się do ogólnych zdolności mentalizacyjnych, emocji czy przywiązania, a także tematów związanych z zaburzeniem osobowości borderline, z tego względu do pracy z osobami osadzonymi wykorzystujemy tylko sześć pierwszych sesji (Zajenkowska i in., 2021). Każdym spotkaniem kieruje lider, który jest odpowiedzialny za grupę przez cały czas jej trwania. Lider koordynuje przebieg spotkania, wprowadza kolejne aktywności, stymuluje dyskusję, w końcu modeluje postawę mentalizującą, zachowując przy tym psychoedukacyjną postawę.

[1] Szczegółowy opis i wskazówki dotyczące programu można znaleźć na stronie: https://www.annafreud.org/media/7863/quality-manual-2018.pdf.

W dalszej części artykułu opisano przebieg poszczególnych spotkań z uwzględnieniem doświadczeń liderów prowadzących warsztaty w różnych grupach, w tym wśród osób osadzonych.

Spotkanie 1: Celem pierwszego warsztatu jest zapoznanie uczestników z terminem „mentalizacja", umożliwienie zrozumienia, czym ona jest i jakie korzyści płyną z prawidłowego i pełnego mentalizowania. Spotkanie rozpoczyna się od przedstawienia się lidera i uczestników, określenia celu programu, planu działania, ustalenia ogólnych zasad. Co ważne, na początku warsztatu lider podkreśla, że spotkania mają charakter edukacyjny, a osobiste trudności uczestników nie będą dogłębnie omawiane i rozwiązywane. Jeżeli uczestnicy zdecydują się na podawanie przykładów z własnego życia, grupa będzie się im przyglądać, zastanawiając się, czego może się z nich nauczyć w kontekście mentalizacji. Przechodząc do właściwego tematu lider wyjaśnia znaczenie terminu „mentalizacja", podkreśla, że w mentalizowaniu „nie ma nic tajemniczego", że jest to prosta koncepcja i coś, co robimy przez większość czasu.

Pierwszą aktywnością grupową jest zastanawianie się nad hipotetyczną sytuacją: lider grupy przedstawia opis scenki, w której występuje mężczyzna stojący na środku skrzyżowania, trzymający w ręku mapę i rozglądający się po okolicy z pytającym wyrazem twarzy. Następnie lider grupy pyta uczestników, co by pomyśleli, gdyby zobaczyli tego mężczyznę. To proste ćwiczenie ma na celu zilustrować, że ludzie interpretują to samo wydarzenie na różne sposoby i że niektóre interpretacje mają charakter mentalizacyjny (np. taka, że mężczyzna jest zagubiony, przestraszony, a inne nie, np. stwierdzenie, że mężczyzna jest turystą – to ważny fakt, który może pomóc w interpretacji, ale nie ma jeszcze charakteru mentalizacyjnego).

Prowadząc warsztaty, np. z nastoletnimi chłopcami, zarówno przebywającymi w zakładzie poprawczym, jak i poza nim, ten moment wymaga szczególnej uważności. Chłopcy bowiem szybko przechodzą do zastanawiania się, co zrobiliby w tej sytuacji (czy pomogliby tej sobie, co by jej powiedzieli, o co zapytali), pomijając to, co dzieje się przed podjęciem decyzji o konkretnym zachowaniu – interpretację. Jest to więc moment na wskazanie przestrzeni dla mentalizacji w codziennych sytuacjach.

W kolejnej części spotkania omawiane są sytuacje, w których mentalizacja jest szczególnie ważna, np. kiedy pocieszamy kogoś bliskiego, chcemy uspokoić płaczące dziecko, jesteśmy na randce, wyjaśniamy nieporozumienie. Uczestnicy podają własne przykłady, a lider podsumowuje, wymieniając kilka powodów, dlaczego mentalizacja ma tak duże znaczenie w relacjach z innymi i samym sobą. Następnie omawiane są różne bieguny mentalizacji, lider objaśnia różnice pomiędzy nimi, podkreśla znaczenie równowagi, podaje przykłady ilustrujące każdy z nich i prosi o przykłady uczestników, by następnie zasugerować, że niepełna mentalizacja może prowadzić do nieporozumień.

W dalszej części toczy się rozmowa na temat tego, dlaczego czasem źle rozumiemy samych siebie i innych, a uczestnicy są proszeni o podanie przykładów takich sytuacji. Na koniec lider grupy wprowadza i objaśnia pojęcie postawy mentalizującej, charakteryzującej się ciekawością doświadczeń, myśli i uczuć drugiej osoby. W celu zilustrowania i lepszego zrozumienia postawy mentalizującej dwóch uczestników zostaje zaproszonych do odgrywania ról – jedna z nich prowadzi wywiad, a jej zadanie polega na tym, aby zastosować postawę mentalizującą i dowiedzieć się, jak druga osoba czuła się wczoraj po południu. Autorzy programu wskazują, że na tym etapie uczestnicy mogą nie czuć się jeszcze w pełni bezpiecznie w grupie, dlatego osobą prowadzącą wywiad

może być lider. Jednakże zazwyczaj pojawiają się ochotnicy (zwykle osoby najbardziej aktywne), którzy wyrażają chęć udziału w ćwiczeniu.

Spotkanie 2: Każde kolejne spotkanie rozpoczyna się od podsumowania tego, co już zostało powiedziane, jest to także czas na wyjaśnienie kwestii, o które zapytają uczestnicy (lub zapewnienie, że kwestie te zostaną poruszone podczas dalszej części warsztatu). Podczas drugiego spotkania lider może starać się motywować do większego zaangażowania osoby, które do tej pory były mało aktywne.

Celem drugiego spotkania jest próba odpowiedzi na pytanie: co to znaczy mieć problemy z mentalizacją? Warsztat rozpoczyna się od ćwiczenia: tym razem lider przedstawia sytuację, w której występuje dwóch bohaterów: dziewczyna i chłopak będący parą. Dziewczyna ma urodziny, więc przygotowuje kolację i czeka na swojego chłopaka. Gdy ten się pojawia, nie ma ze sobą prezentu, za to dziwi się, że jego partnerka przygotowała taką wykwintną kolację i to we wtorek. Podczas kolacji dziewczyna jest spokojna, ale wypija całą butelkę wina. Zadaniem uczestników jest zastanowienie się, co się stało i dlaczego dziewczyna zachowała się w taki sposób. Uczestnicy zwykle szybko dochodzą do wniosku, że chłopak zapomniał o urodzinach i dziewczyna jest na niego zła, trudniej im jednak odnieść stan psychiczny do zachowania związanego z wypiciem dużej ilości alkoholu. Ćwiczenie to jest rozgrzewką przed wprowadzeniem tematu słabych zdolności mentalizacyjnych, charakteryzujących się m.in.: myśleniem w kategoriach czarno-białych, poczuciem zbytniej pewności co do motywów innych ludzi czy małą ciekawością stanów psychicznych.

Następnie lider grupy pyta uczestników o możliwe konsekwencje słabej mentalizacji w stosunku do innych oraz w stosunku do samych siebie. Po dyskusji na ten temat prowadzący podsumowuje, że problemy z mentalizacją mogą doprowadzić do trudności w relacjach z innymi ludźmi, słabej kontroli emocjonalnej, impulsywności czy niestabilnego poczucia własnej wartości. W kolejnej części program płynnie przechodzi do tematu emocji, lider wyprzedza nieco program i wskazuje uczestnikom, że jedną z najważniejszych przyczyn słabej mentalizacji jest silna aktywacja emocjonalna, następnie prosi, aby uczestnicy zastanowili się, co dzieje się z nimi, jaki jest typowy wzorzec ich reakcji, kiedy pojawiają się silne emocje. Podczas warsztatów wśród osób osadzonych na tym etapie często pojawia się myśl, że uczestnicy są osobami bardzo spokojnymi i nieczęsto odczuwają emocje.

W ostatniej części spotkania osoba prowadząca prezentuje wykres pokazujący zależność pomiędzy zdolnością do mentalizacji a pobudzeniem/stresem. Do pewnego momentu związek ten jest współliniowy, jednak po przekroczeniu progu pobudzenia tracimy zdolność do mentalizowania w sposób świadomy i zaczynamy mentalizować wyłącznie w sposób automatyczny, co utrudnia, a wręcz uniemożliwia kontrolowanie własnych myśli i zachowań. Na bazie tego schematu lider wyjaśnia, że ludzie różnią się pod względem tego, jak szybko aktywowane są u nich emocje, oraz że w różnym momencie przekraczają próg pobudzenia i potrzebują różnej ilości czasu na powrót do równowagi. Na koniec tej sesji uczestnicy są proszeni o przyjrzenie się przykładom sytuacji, w których ich zdolność do mentalizowania została osłabiona.

Spotkanie 3: Celem tego spotkania jest omówienie podstawowych rodzajów emocji i roli, jaką pełnią w naszym życiu. Wprowadzając do tematu, lider pyta uczestników o to, jakie znają emocje. Podczas warsztatów często w tym momencie pojawia się dyskusja na temat emocji pozytywnych i negatywnych, zwykle w tym momencie pro-

wadzący sugeruje, że trudno dokonać takiego jednoznacznego podziału, bo wszystkie emocje są nam potrzebne, i proponuje inną kategoryzację, na emocje podstawowe i złożone. Lider wyjaśnia, krótko, czym charakteryzują się te dwie kategorie, i prosi, aby uczestnicy zasugerowali, które emocje zaliczyliby do tych podstawowych. Podczas ich omawiania lider posługuje się teorią Panksеppa (1998), który zalicza do grupy emocji podstawowych siedem programów działania: poszukiwanie, wściekłość, lęk, pożądanie, opiekę, smutek i zabawę. Następnie każda z tych emocji jest omawiana w kontekście jej głównych zadań. Lider podkreśla, że ludzie mają zdolność tłumienia emocji, dlatego czasem pojawiają się uczucia, które są świadomym doświadczeniem stanu ciała podczas aktywności emocjonalnej, ale uczuciom tym nie towarzyszy wiedza na temat specyficznego charakteru emocji.

Po części teoretycznej jest czas na dyskusję, dotyczącą tego, czy wszyscy w grupie odczuwają różne emocje, czy każda osoba doświadcza ich z taką samą częstotliwością i intensywnością. Podczas pracy z osobami osadzonymi dyskusja ta zwykle okazywała się znacząca. Na jednym z warsztatów prowadzonym pośród dorosłych przebywających w zakładzie karnym grupa zgodziła się co do tego, że w więzieniu nie ma smutku, po chwili dyskusji jednak pojawiała się myśl podważająca tę tezę – w więzieniu nie okazuje się smutku, bo można być uznanym za słabszego od innych, ale to nie znaczy, że osadzeni go nie odczuwają.

Spotkanie 4: Kontynuowany jest temat związany z emocjami. Spotkanie rozpoczyna się od krótkiego podsumowania poprzedniego warsztatu, następnie lider zachęca do dyskusji: w jaki sposób rozpoznajemy emocje – innych, a następnie nas samych. Dotychczasowe warsztaty pokazały, że uczestnikom łatwiej jest mówić o emocjach innych ludzi niż własnych. Zwykle szybko podają możliwe sposoby rejestrowania emocji: interpretacje, np. mimiki, gestów, postawy ciała, sposobu poruszania, rzadziej mówią o identyfikacji, która jest możliwa dzięki neuronom lustrzanym i jest jedną ze składowych empatii. Dyskusja o własnych emocjach jest mniej dynamiczna i trudniejsza, dlatego też lider wprowadza ćwiczenie grupowe: prosi uczestników, aby zamknęli oczy, zapomnieli o otoczeniu i skupili się na sobie. Prowadzący kieruje uwagę uczestników do wewnątrz, zadając pytania, np. czy jest jakieś miejsce w Twoim ciele, które przyciąga uwagę? Uczestnicy zwykle z rezerwą podchodzą do propozycji tego ćwiczenia, ostatecznie jednak zgadzają się wziąć w nim udział. Po jego zakończeniu dyskusja o własnych emocjach zdecydowanie ożywa.

Kolejnym etapem sesji jest rozmowa o możliwościach regulacji emocji przez inne osoby – ćwiczenie dotyczące pocieszania, tego, jakie zachowania innych osób mogą być pomocne, pocieszające. W końcu rozmowa o tym, co uczestnicy robią, aby poradzić sobie z nieprzyjemnymi stanami emocjonalnymi, pojawia się także zadanie domowe, które polega na zanotowaniu przykładów sytuacji, kiedy uczestnikom udało się uregulować nieprzyjemny stan emocjonalny. Jak pokazują badania, jest to ważne zagadnienie, ponieważ osoby przejawiające trudności w regulacji emocji mają dłuższą historię kryminalną; autorzy sugerują też, że programy przeznaczone dla osadzonych powinny skupiać się raczej na kontrolowaniu zachowań w sytuacjach, kiedy pojawia się gniew, niż na jego tłumieniu (Roberton i in., 2014). Proponowany program spełnia założenia tego postulatu.

Spotkanie 5: Na przedostatnim spotkaniu jest omawiany temat przywiązania, które lider definiuje jako: „pozytywne uczucie i emocjonalną więź z drugim człowiekiem". Ten

warsztat rozpoczyna się od wykładu na temat pierwszych relacji z opiekunami, wzorców przywiązania, sposobów ich badania oraz znaczenia, jakie mają dla budowania kolejnych relacji. Po części teoretycznej prowadzący ponownie przytacza opis pewnej sytuacji. Po raz drugi opowiada o chłopaku i dziewczynie, spotykających się ponownie po wakacjach, podczas których chłopak ani razu nie zadzwonił do dziewczyny, a kiedy ona dzwoniła, nie odbierał. Dziewczyna nie miała zbyt aktywnych wakacji, ale kiedy chłopak zapytał ją, jak minęły, odpowiedziała, że świetnie się bawiła i chciałaby, aby trwały dłużej. Zadaniem uczestników jest przedyskutowanie tej historii w kontekście przywiązania i zastanowienie się nad odpowiedzią na pytanie: dlaczego dziewczyna odpowiedziała w taki sposób? Ćwiczenie to pozwala liderowi sprawdzić, na ile omawiane zagadnienie jest zrozumiałe dla uczestników, i przejść do kolejnego, trudniejszego ćwiczenia: dyskusji na temat własnych relacji. Jest to bardzo wymagająca część warsztatu, dlatego nie w każdej grupie w tym momencie rozwija się dyskusja.

Spotkanie 6: Ostatni warsztat dotyczy przywiązania i mentalizacji. Lider podsumowuje najważniejsze informacje dotyczące tych konstruktów z poprzednich spotkań, by następnie opowiedzieć o środowisku sprzyjającym mentalizacji: jest to takie środowisko, w którym rozmawia się na temat tego, dlaczego ludzie zachowują się w określony sposób, co mogą czuć, jakie mają motywacje. Dorastanie w takim środowisku sprzyja bezpiecznym relacjom i pozwala na rozwój zdolności mentalizowania. Podczas aktywności grupowej uczestnicy są proszeni o zastanowienie się, co charakteryzuje ich środowisko rodzinne w odniesieniu do mentalizacji? Wśród osadzonych dyskusja schodzi często w tym momencie na temat środowiska więziennego.

Kolejnym zagadnieniem poruszanym na ostatnim spotkaniu jest konflikt związany z przywiązaniem, który pojawia się wówczas, „kiedy wyolbrzymiamy lub hamujemy reakcje emocjonalne, ponieważ obawiamy się, co się stanie, jak zaalarmujemy drugą osobę". Grupa jest proszona o rozważenie własnych przykładów takich konfliktów, a lider wskazuje na ich możliwe konsekwencje dla rozwoju zdolności mentalizacyjnych. Podczas kolejnej aktywności uczestnicy zastanawiają się nad tematami, o których trudno im mówić w bliskim związku.

W ostatniej części tego spotkania jest czas na podsumowanie, zadanie dodatkowych pytań, podzielenie się wrażeniami i wyjście z trudnych tematów. Jeden z uczestników – młody mężczyzna – podzielił się pewnego razu refleksją, że warsztat dotyczący mentalizacji jest jak warsztat dotyczący oddychania, wszyscy oddychamy i mentalizujemy na co dzień, tylko w przypadku mentalizacji zwykle nie znamy nazwy.

Redukowanie wrogości i konkluzje

Rozwijanie mentalizacji poprawia zdolności interpersonalne i usprawnia funkcjonowanie społeczne, co zostało wykazane w wielu badaniach empirycznych (np. Bateman, Fonagy, 1999, 2008, 2009, 2010, 2019; Robinson i in., 2014), głównie z udziałem pacjentów. Część psychoedukacyjna programu MBT wykazuje także obiecujące efekty w redukowaniu poziomu wrogich interpretacji dokonywanych wobec innych osób, co zostało zbadane wśród osadzonych kobiet (Zajenkowska i in., 2021). Autorzy badania przeprowadzili jakościową analizę wypowiedzi uczestniczek podczas trwania warsztatów, przeanalizowali także odpowiedzi udzielane w kwestionariuszu AIHQ (Ambiguous Intentions Hostility Questionnaire, Combs i in., 2007, Zajenkowska i in., 2018).

Narzędzie to zawiera opisy pięciu sytuacji (np. *Pracujesz w nowym miejscu od trzech tygodni. Pewnego dnia widzisz jednego ze swoich kolegów/jedną z koleżanek z pracy na ulicy. Idziesz w jego/jej stronę, aby się przywitać, ale on/ona przechodzi koło Ciebie, nie odzywając się*), zadaniem osoby badanej jest odpowiedzenie na kilka pytań, zarówno otwartych (np. *Jak myślisz, dlaczego Twój kolega/Twoja koleżanka z pracy minął/-ęła Cię bez przywitania? Co byś zrobił/-ła w tej sytuacji?*), jak i zamkniętych (np. *Czy myślisz, że Twój kolega/Twoja koleżanka zrobił/-ła to celowo?*). Kwestionariusz był wypełniany dwukrotnie – przed warsztatem i po nim. Już po sześciu sesjach uczestniczki zaczęły przyjmować perspektywę innych ludzi (np. nielubianych strażników), ponadto były bardziej zaangażowane w tworzenie alternatywnych interpretacji na temat zachowań, podejmowały także próbę zrozumienia intencji innych osób (Zajenkowska i in., 2021). Obecnie trwają badania w innych grupach, także wśród osadzonych mężczyzn i chłopców. W fazie testów jest również skuteczność warsztatu prowadzonego w formie online (APSychoLab). Wyniki wydają się obiecujące, coraz więcej ośrodków penitencjarnych wyraża także zainteresowanie współpracą.

Opisany program ma jednak pewne ograniczenia: przede wszystkim wymaga ścisłej współpracy badaczy, trenerów i psychologów pracujących w instytucjach, gdzie prowadzony jest warsztat. Wśród osób przebywających w izolacji mówienie o emocjach, mentalizacji czy przywiązaniu może zwiększyć nasilenie objawów depresyjnych (np. Roth, Cohen, 1986), dlatego osadzeni po zakończeniu warsztatu powinni być objęci opieką psychologiczną. Warsztaty są także wyzwaniem organizacyjnym, przede wszystkim dla służby więziennej, która ma dodatkowe obowiązki związane z przyprowadzaniem i odprowadzaniem osadzonych czy zapewnieniem bezpieczeństwa trenerom (jeżeli warsztat odbywa się w tradycyjnej formie).

Wydaje się jednak, że warto podejmować ten wysiłek. Wiedza na temat mentalizacji i wyniki dotychczasowych badań pozwalają na stwierdzenie, że rozwijanie umiejętności mentalizowania – rozumienia samego siebie i innych ludzi – może zwiększyć elastyczność i otwartość w odbieraniu oraz interpretacji informacji o świecie społecznym, zmniejszyć tendencje do przypisywania wrogich atrybucji w niejednoznacznych sytuacjach, a tym samym zredukować prawdopodobieństwo wystąpienia agresji. Warsztaty psychoedukacyjne z elementami mentalizacji zwiększają szanse, że kiedy osoba, która wzięła w nich udział, znajdzie się w hipotetycznej sytuacji, opisanej na początku (usłyszy, jak inne osoby zaczynają się śmiać), będzie dysponowała zasobami pozwalającymi na stworzenie wielu różnych interpretacji tej sytuacji i zamiast atakować, zapyta albo przejdzie obojętnie.

BIBLIOGRAFIA

Adamczyk, L. (2013a). Mentalizacja. Cz. 1: Wprowadzenie do zagadnienia, wymiary mentalizacji. *Psychoterapia*, *3*(166), 25–36.

Adamczyk, L. (2013b). Mentalizacja. Cz. 2: Neurofizjologiczne podłoże mentalizacji, pojęcia zbliżone zakresem i znaczeniem do mentalizacji, przedmentalizacyjne tryby funkcjonowania psychicznego, ocena mentalizacji. *Psychoterapia*, *3*(166), 37–46.

Asen, E., Fonagy, P. (2011). Mentalization-based family therapy [Terapia rodzinna oparta na mentalizacji]. W: Bateman, A., Fonagy, P. (red.), *Handbook of mentalizing in mental health practice* [Podręcznik mentalizacji w praktyce zdrowia psychicznego] (s. 107–128). American Psychiatric Publishing.

Bateman, A., Fonagy, P. (1999). Effectiveness of partial hospitalization in the treatment of borderline personality disorder:

a randomized controlled trial. *American Journal of Psychiatry*, *156*(10), 1563–1569.

Bateman, A., Fonagy, P. (2001). Treatment of borderline personality disorder with psychoanalytically oriented partial hospitalisation: an 18-month follow-up. *The American Journal of Psychiatry*, *158*, 36–42. https://doi.org/10.1176/appi.ajp.158.1.36

Bateman, A., Fonagy, P. (2004). *Psychotherapy for borderline personality disorder* [Psychoterapia zaburzeń osobowości typu borderline]. Oxford University Press.

Bateman, A., Fonagy, P. (2008). 8-year follow-up of patients treated for borderline personality disorder: mentalization-based treatment versus treatment as usual. *American Journal of Psychiatry*, *165*(5), 631–638. https://doi.org/10.1176/appi.ajp.2007.07040636

Bateman, A., Fonagy, P. (2009). Randomized controlled trial of outpatient mentalization-based treatment versus structured clinical management for borderline personality disorder. *American Journal of Psychiatry*, *166*(12), 1355–1364. https://doi.org/10.1176/appi.ajp.2009.09040539

Bateman, A., Fonagy, P. (2010). Mentalization based treatment for borderline personality disorder. *World Psychiatry*, *9*(1), 11–15. https://doi.org/10.1002/j.2051-5545.2010.tb00255.x

Bateman, A., Fonagy, P. (2011). *Handbook of mentalizing in mental health practice* [Podręcznik mentalizacji w praktyce zdrowia psychicznego]. American Psychiatric Publishing.

Bateman, A., Fonagy, P. (2013). Mentalization-based treatment. *Psychoanalytic Inquiry*, *33*(6), 595–613. https://doi.org/10.1080/07351690.2013.835170

Bateman, A., Fonagy, P. (2019). Mentalization-based treatment for borderline and antisocial personality disorder. W: *Contemporary psychodynamic psychotherapy* (s. 133–148). Academic Press.

Beer, J. S., Ochsner, K. N. (2006). Social cognition: a multi level analysis. *Brain Research*, *1079*(1), 98–105. https://doi.org/0.1016/j.brainres.2006.01.002

Bowlby, J. (1973). *Attachment and loss: Volume II: Separation, anxiety and anger.* The Hogarth Press and the Institute of Psychoanalysis.

Bowlby, J. (2007). *Przywiązanie*. Wydawnictwo Naukowe PWN.

Caspi, A. (2000). The child is father of the man: personality continuities from childhood to adulthood. *Journal of Personality and Social Psychology*, *78*(1), 158–172. https://doi.org/10.1037/0022-3514.78.1.158

Choi-Kain, L. W., Gunderson, J. G. (2008). Mentalization: ontogeny, assessment, and application in the treatment of borderline personality disorder. *The American Journal of Psychiatry*, *165*(9), 1127–1135. https://doi.org/10.1176/appi.ajp.2008.07081360

Cierpiałkowska, L., Górska, D. (2016). Psychologia zaburzeń osobowości. W: L. Cierpiałkowska, H. Sęk (red.), *Psychologia kliniczna* (s. 283–303). Wydawnictwo Naukowe PWN.

Coates, S. (2006). Foreword [Przedmowa]. W: J. G. Allen, P. Fonagy (red.), *Handbook of mentalizing in mental health practice* [Podręcznik mentalizacji w praktyce zdrowia psychicznego] (s. 15–18). John Wiley & Sons Ltd.

Combs, D. R., Penn, D. L., Wicher, M., Waldheter, E. (2007). The Ambiguous Intentions Hostility Questionnaire (AIHQ): a new measure for evaluating hostile social-cognitive biases in paranoia. *Cognitive Neuropsychiatry*, *12*(2), 128–143. https://doi.org/10.1080/13546800600787854

Crain, M. M., Finch, C. L., Foster, S. L. (2005). The relevance of the social information processing model for understanding relational aggression in girls. *Merrill-Palmer Quarterly* (1982), 213–249.

Crick, N. R., Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin*, *115*(1), 74–101. https://doi.org/10.1037/0033-2909.115.1.74

Crick, N. R., Dodge, K. A. (1996). Social information-processing mechanisms in reactive and proactive aggression. *Child Development*, *67*(3), 993–1002. https://doi.org/10.2307/1131875

Dodge, K. A. (1980). Social cognition and children's aggressive behavior. *Child De-*

velopment, *51*(1), 162–170. https://doi.org/10.2307/1129603

Dodge, K. A. (1986). A social information processing model of social competence in children [Model przetwarzania informacji społecznych kompetencji społecznych u dzieci]. W: M. Perlmutter (red.), *The Minnesota Symposium on Child Psychology* [Sympozjum w Minnesocie na temat psychologii dziecka] (t. 18, s. 77–125). Hillsdale, NJ: Erlbaum.

Dodge, K. A. (2006). Translational science in action: Hostile attributional style and the development of aggressive behavior problems. *Development and Psychopathology*, *18*(3), 791–814.

Dykas, M. J., Cassidy, J. (2011). Attachment and the processing of social information across the life span: theory and evidence. *Psychological Bulletin*, *137*(1), 19–46. https://doi.org/10.1037/a0021367

Epps, J., Kendall, P. C. (1995). Hostile attributional bias in adults. *Cognitive Therapy and Research*, *19*(2), 159–178. https://doi.org/10.1007/BF02229692

Farrington, D. P., Jolliffe, D., Loeber, R., Stouthamer-Loeber, M., Kalb, L. M. (2001). The concentration of offenders in families, and family criminality in the prediction of boys' delinquency. *Journal of Adolescence*, *24*(5), 579–596. https://doi.org/10.1006/jado.2001.0424

Fonagy, P. (1991). Thinking about thinking: Some clinical and theoretical considerations in the treatment of a borderline patient. *International Journal of Psychoanalysis*, *72*, 1–18.

Fonagy, P. (2003). The development of psychopathology from infancy to adulthood: The mysterious unfolding of disturbance in time. *Infant Mental Health Journal: Official Publication of The World Association for Infant Mental Health*, *24*(3), 212–239. https://doi.org/10.1002/imhj.10053

Fonagy, P., Levinson, A. (2004). Offending and attachment: The relationship between interpersonal awareness and offending in a prison population with psychiatric disorder. *Canadian Journal of Psychoanalysis*, *12*(2), 225–251.

Fonagy, P., Bateman, A. W., Luyten, P. (2011). Introduction and overview [Wprowadzenie i przegląd]. W: A. W. Bateman, P. Fonagy (red.), *Handbook of mentalizing in mental health practice* [Podręcznik mentalizacji w praktyce zdrowia psychicznego] (s. 3–42). American Psychiatric Publishing.

Fonagy, P., Gergely, G., Jurist, L. E., Target, M. (2004). *Affect regulation, mentalization, and the development of the self* [Regulacja afektu, mentalizacja i rozwój jaźni]. Karnac.

Gerrig, R. J. (1988). Text comprehension [Zrozumienie tekstu]. W: R. J. Steinberg, E. E. Smith (red.), *The psychology of human thought* [Psychologia ludzkiej myśli] (s. 242–266). Cambridge University Press.

Hay, D. F. (2005). The beginnings of aggression in infancy [Początki agresji w dzieciństwie]. W: R. E. Tremblay, W. W. Hartup, J. Archer (red.), *Developmental origins of aggression* [Początki rozwojowe agresji] (s. 107–132). Guilford Press.

Horsley, T. A., de Castro, B. O., van der Schoot, M. (2010). In the eye of the beholder: Eye-tracking assessment of social information processing in aggressive behavior. *Journal of Abnormal Child Psychology*, *38*(5), 587–599. https://doi.org/10.1007/s10802-009-9361-x

Luyten, P., Campbell, C., Allison, E., Fonagy, P. (2020). The mentalizing approach to psychopathology: State of the art and future directions. *Annual Review of Clinical Psychology*, *16*, 297–325. https://doi.org/10.1146/annurev-clinpsy-071919-015355

Luyten, P., Fonagy, P., Lowyck, B., Vermote, R. (2012). Assessment of mentalization [Ocena mentalizacji]. W: A. W. Bateman, P. Fonagy (red.), *Handbook of mentalizing in mental health practice* [Podręcznik mentalizacji w praktyce zdrowia psychicznego] (s. 43–65). American Psychiatric Publishing.

Ogilvie, C. A., Newman, E., Todd, L., Peck, D. (2014). Attachment and violent offending: A meta-analysis. *Aggression and Violent Behavior*, *19*(4), 322–339. https://doi.org/10.1016/j.avb.2014.04.007

Papaleontiou-Louca, E. (2008). *Metacognition and theory of mind* [Metapoznanie

i teoria umysłu]. Newcastle: Cambridge Scholars Publishing.

Roberton, T., Daffern, M., Bucks, R. S. (2014). Maladaptive emotion regulation and aggression in adult offenders. *Psychology, Crime i Law*, *20*(10), 933–954. https://doi.org/10.1186/1471-244X-14-51

Robinson, P., Barrett, B., Bateman, A., Hakeem, A., Hellier, J., Lemonsky, F., Rutterford, C., Schmidt, U., Fonagy, P. (2014). Study protocol for a randomized controlled trial of mentalization based therapy against specialist supportive clinical management in patients with both eating disorders and symptoms of borderline personality disorder. *BMC Psychiatry*, *14*(1), 1–10. https://doi.org/0.1186/1471-244X-14-51

Roth, S., Cohen, L. J. (1986). Approach, avoidance, and coping with stress. *American Psychologist*, *41*(7), 813–819. https://doi.org/10.1037/0003-066X.41.7.813

Spielberger, C. D., Krasner, S. S., Solomon, E. P. (1988). The experience, expression, and control of anger [Doświadczenie, ekspresja i kontrola gniewu]. W: M. P. Janisse (red.), *Individual differences, stress, and health psychology* [Różnice indywidualne, stres i psychologia zdrowia] (s. 89–108). Springer.

Suter, J. M., Byrne, M. K., Byrne, S., Howells, K., Day, A. (2002). Anger in prisoners: Women are different from men. *Personality and Individual Differences*, *32*(6), 1087–1100. https://doi.org/10.1016/S0191-8869(01)00105-2

Volkert, J., Hauschild, S., Taubner, S. (2019). Mentalization-based treatment for personality disorders: efficacy, effectiveness, and new developments. *Current Psychiatry Reports*, *21*(4), 1–12. https://doi.org/10.1007/s11920-019-1012-5

Zajenkowska, A., Prusik, M., Szulawski, M. (2018). What does the Ambiguous Intentions Hostility Questionnaire really measure? The importance of context in evaluating hostility bias. *Journal of Personality Assessment*, *102*(2), 205–213. https://doi.org/10.1080/00223891.2018.1525389

Zajenkowska, A., Prusik, M., Jasielska, D., Szulawski, M. (2021). Hostile attribution bias among offenders and non-offenders: Making social information processing more adequate. *Journal of Community & Applied Social Psychology*, *31*(2), 241–256. https://doi.org/10.1002/casp.2493

Zajenkowska, A., Rajchert, J. (2020). How sensitivity to provocation shapes encoding and interpretation of ambivalent scenes in an eye tracking study. *Journal of Cognitive Psychology*, *32*(2), 180–198. https://doi.org/10.1080/20445911.2020.1717498

Zajenkowska, A., Rajchert, J., Macianowicz, O., Holas, P., Murawiec, S. (2019). Cognitive-Behavioral (CBT) and Psychodynamic (PDT) group psychotherapy and their impact on depressive symptoms and hostile attributions. *International Journal of Group Psychotherapy*, *69*(4), 383–407. https://doi.org/10.1080/00207284.2019.1653189

PSYCHO-EDUCATIONAL TRAINING WITH ELEMENTS OF MENTALIZATION AND THE ROLE IT PLAYS IN REDUCING HOSTILITY AMONG INMATES

Abstract: According to the Social Information Processing Theory, behavior is the product of six, consecutive steps. The first two – encoding and interpreting – are crucial, as they could determine how the next steps in the process will proceed. Adequate interpretation of one's own and other people's behavior is possible thanks to a developed ability to mentalize. Mentalizing is an imaginative mental activity that aims to realize human behavior as intentional states of mind. Incomplete or abnormal mentalization hinders effective functioning in the social world and may, for example, increase the risk of aggressive behavior. Previous research has shown that imprisoned individuals demonstrate reduced ability to mentalize. At the same time, it is a capacity that can be developed at any stage of life. Therefore, we propose a psycho-educational workshop with elements of mentalization, we describe its course, the experience of the conductors (leaders), and the preliminary research results confirming the effectiveness of the method.

Keywords: mentalization, attachment, inmates, social information processing

Warszawa, 04.11.2022

Imię i Nazwisko: Anna Zajenkowska

Afiliacja: Akademia Pedagogiki Specjalnej

Oświadczenie

A) Oświadczam, że w artykule:

Gehrer, N. A., **Zajenkowska, A.**, Bodecka, M., & Schönenberg, M. (2021). Attention orienting to the eyes in violent female and male offenders: An eye-tracking study. *Biological psychology*, *163*, 108136.

- Mój wkład polegał na: opracowaniu koncepcji badań, nadzorze badań, współprzygotowaniu pierwszej wersji artykułu oraz odpowiedzi na recenzje.

- Wkład Marty Bodeckiej-Zych polegał na: prowadzeniu badań, wstępnym opracowywaniu danych, pomocy przy tworzeniu pierwszej wersji artykułu - wkład procentowy: 10%

Podpis

B) Oświadczam, że w artykule:

Zajenkowska, A. M., Bodecka, M., Duda, E., & Lawrence, C. (2022). Reduced attention toward faces, intentionality and blame ascription in violent offenders and community-based adults: Evidence from an eye-tracking study. *Aggressive Behavior*. *48*(2), 264-275

- Mój wkład polegał na: opracowaniu koncepcji badań, nadzorze badań, współprzygotowaniu pierwszej wersji artykułu oraz odpowiedzi na recenzje.

- Wkład Marty Bodeckiej-Zych polegał na: prowadzeniu badań, wstępnym opracowywaniu danych, pomocy przy tworzeniu pierwszej wersji artykułu i odpowiedzi na recenzje - wkład procentowy: 20%

Podpis

C) Oświadczam, że w artykule:

Bodecka-Zych, M., **Zajenkowska, A**., & Bower Russa, M. (2022). Sex Differences in Inmates: Anger, Sensitivity to Provocation and Family History of Imprisonment. *International Journal of Offender Therapy and Comparative Criminology, 66*(12), 1327-1342.

- Mój wkład polegał na: konsultowaniu całości manuskryptu oraz współprzygotowaniu odpowiedzi na recenzje.

- Wkład Marty Bodeckiej-Zych polegał na: prowadzeniu badań, przeprowadzeniu analiz statystycznych, przygotowaniu pierwszej wersji manuskryptu i odpowiedzi na recenzje - wkład procentowy: 70%

Podpis

D) Oświadczam, że w artykule:

Zajenkowska, A., Bodecka-Zych, M., Gehrer, N., Krejtz, K., Lawrence, C., Schönenberg, M., & Jusyte, A. (2022). Gender differences in sensitivity to provocation and hostile attribution bias toward ambiguous facial cues in violent offenders and community-based adults. *Motivation and Emotion*, 1-10.

- Mój wkład polegał na: opracowaniu koncepcji badań, nadzorze badań, współprzygotowaniu pierwszej wersji artykułu, oraz odpowiedzi na recenzje

- Wkład Marty Bodeckiej-Zych polegał na: prowadzeniu badań, wstępnym opracowywaniu danych, pomocy przy tworzeniu pierwszej wersji artykułu i odpowiedzi na recenzje - wkład procentowy: 20%

Podpis

E) Oświadczam, że w artykule:

Bodecka-Zych, M., Jonason, P. K., & **Zajenkowska, A.** (2022). Hostile attribution biases in vulnerable narcissists depends on the socio-relational context. *Journal of Individual Differences, 43*(2), 70-78.

- Mój wkład polegał na: konsultowaniu całości manuskryptu oraz testowanego modelu.

- Wkład Marty Bodeckiej-Zych polegał na: prowadzeniu badań, przeprowadzeniu analiz statystycznych, przygotowaniu pierwszej wersji manuskryptu i odpowiedzi na recenzje - wkład procentowy: 70%

Podpis

F) Oświadczam, że w artykule:

Bodecka-Zych, M., **Zajenkowska, A**., & Lawrence, C. (2022). Dad, are they laughing at me? Fathers' vulnerable narcissism and sons' hostile attributions. *Personality and Individual Differences, 192*, 111582.

- Mój wkład polegał na: konsultowaniu całości manuskryptu oraz testowanego modelu.

- Wkład Marty Bodeckiej-Zych polegał na: konceptualizacji i zaprojektowaniu badania, prowadzeniu badań, przeprowadzeniu analiz statystycznych, przygotowaniu pierwszej wersji manuskryptu i odpowiedzi na recenzje - wkład procentowy: 80%

Podpis

G) Oświadczam, że w artykule:

Bodecka, M., Jakubowska, A., **Zajenkowska, A.** (2021). Warsztat psychoedukacyjny z elementami mentalizacji i jego rola w redukowaniu wrogości wśród osób osadzonych.[Psycho-educational workshop with elements of mentalization and its role in reducing hostility among inmates] *Psychologia Wychowawcza, 62*(20), 101-115.

- Mój wkład polegał na: konsultowaniu całości manuskryptu oraz współprzygotowaniu odpowiedzi na recenzje.

- Wkład Marty Bodeckiej-Zych polegał na: prowadzeniu warsztatów, przygotowaniu pierwszej wersji manuskryptu i odpowiedzi na recenzje – wkład procentowy: 80%

Podpis

Tübingen, 11/11/2022

Name: Dr. Nina Gehrer

Affiliation: University of Tübingen

Statement

A) I declare that in the paper:

Gehrer, N. A., Zajenkowska, A., Bodecka, M., & Schönenberg, M. (2021). Attention orienting to the eyes in violent female and male offenders: An eye-tracking study. *Biological psychology*, *163*, 108136.

- My contribution consisted of the following: Conceptualization; Methodology; Data Analysis; Validation; Writing - Original Draft; Writing - Review & Editing
- Marta Bodecka-Zych's contribution consisted of: conducting research, preliminary data processing, assistance with the first version of the article - percentage contribution: 10%

B) I declare that in the paper:

Zajenkowska, A., Bodecka-Zych, M., **Gehrer, N**., Krejtz, K., Lawrence, C., Schönenberg, M., & Jusyte, A. (2022). Gender differences in sensitivity to provocation and hostile attribution bias toward ambiguous facial cues in violent offenders and community-based adults. *Motivation and Emotion*, 1-10.

- My contribution consisted of the following: Methodology; Writing - Review & Editing
- Marta Bodecka-Zych's contribution consisted of: conducting the research, preliminary data processing, assistance with the first version of the article and responding to reviews - percentage contribution: 20%

Tübingen, 11/11/2022

Name: Dr. Schönenberg

Affiliation: University of Tübingen

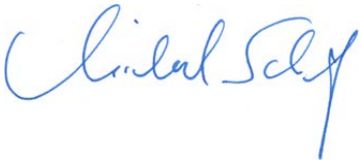
Statement

A) I declare that in the paper:

Gehrer, N. A., Zajenkowska, A., Bodecka, M., & **Schönenberg, M**. (2021). Attention orienting to the eyes in violent female and male offenders: An eye-tracking study. *Biological Psychology, 163*, 108136.

- My contribution consisted of the following: Conceptualization; Methodology; Validation; Writing - Original Draft; Writing - Review & Editing; Supervision

- Marta Bodecka-Zych's contribution consisted of: conducting research, preliminary data processing, assistance with the first version of the article - percentage contribution: 10%

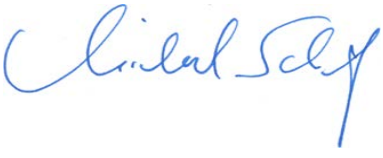


B) I declare that in the paper:

Zajenkowska, A., Bodecka-Zych, M., Gehrer, N., Krejtz, K., Lawrence, C., **Schönenberg, M.**, & Jusyte, A. (2022). Gender differences in sensitivity to provocation and hostile attribution bias toward ambiguous facial cues in violent offenders and community-based adults. *Motivation and Emotion*, 1-10.

- My contribution consisted of the following: Methodology; Writing - Original Draft; Writing - Review & Editing; Supervision

- Marta Bodecka-Zych's contribution consisted of: conducting the research, preliminary data processing, assistance with the first version of the article and responding to reviews - percentage contribution: 20%



Warszawa, 10.11.2022

Ewa Duda:

Akademia Pedagogiki Specjalnej

im. Marii Grzegorzewskiej

Oświadczenie

Oświadczam, że w artykule:

Zajenkowska, A. M., Bodecka, M., **Duda, E.**, & Lawrence, C. (2022). Reduced attention toward faces, intentionality and blame ascription in violent offenders and community-based adults: Evidence from an eye-tracking study. *Aggressive Behavior. 48*(2), 264-275

- Mój wkład polegał na: opracowaniu danych, przeprowadzeniu analizy danych, tworzeniu treści artykułu (opis grupy badawczej, część dotycząca wyników) i współpracy przy odpowiedzi na recenzje.

- Wkład Marty Bodeckiej-Zych polegał na: prowadzeniu badań, wstępnym opracowywaniu danych, pomocy przy tworzeniu pierwszej wersji artykułu i odpowiedzi na recenzje - wkład procentowy: 20%

Podpis

Ewa Duda

Allendale, Mi., 11/16/2022

Name: Mary Bower Russa

Affiliation: Grand Valley State University (Allendale, Michigan)

Statement

I declare that in the paper:

Bodecka-Zych, M., Zajenkowska, A., & **Bower Russa, M.** (2022). Sex Differences in Inmates: Anger, Sensitivity to Provocation and Family History of Imprisonment. *International Journal of Offender Therapy and Comparative Criminology, 66*(12), 1327-1342.

- My contribution consisted of the following: I assisted with the preparation of the first version of the article, and the response to reviewers. In this role I provided suggestions regarding some relevant literature, ways to possibly frame some of the findings, and I assisted with some editing.
- Marta Bodecka-Zych's contribution consisted of: conducting the research, conducting statistical analyses, preparing the first version of the manuscript and responding to reviews - percentage contribution: 70%

Statement

I declare that in the paper:

Bodecka-Zych, M., **Jonason, P. K**., & Zajenkowska, A. (2022). Hostile attribution biases in vulnerable narcissists depends on the socio-relational context. *Journal of Individual Differences*, *43*(2), 70.

- My contribution consisted of the following: Guidance in statistics and writing only. English-language editing. Advice about where to submit.

- Marta Bodecka-Zych's contribution consisted of: conducting the research, conducting statistical analyses, preparing the first version of the manuscript and responding to reviews - percentage contribution: 70%

PK Jonason, Ph.D.

www.peterjonason.com

Nottingham, 14th November 2022

Name: Claire Lawrence

Affiliation: Lawrence PsychAdvisory (previously Associate Professor, School of Psychology, University of Nottingham Nottingham NG7 2RD, UK)

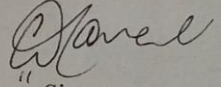
Statement

A) I declare that in the paper:

Zajenkowska, A. M., Bodecka, M., Duda, E., & **Lawrence, C**. (2022). Reduced attention toward faces, intentionality and blame ascription in violent offenders and community-based adults: Evidence from an eye-tracking study. *Aggressive Behavior. 48*(2), 264-275

- My contribution consisted of the following: Comment and guidance on data analysis and earlier drafts.

- Marta Bodecka-Zych's contribution consisted of: conducting the research, preliminary data processing, assistance with the first version of the article and responding to reviews - percentage contribution: 20%

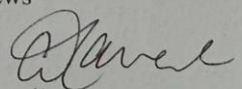


" Signature

B) I declare that in the paper:

Zajenkowska, A., Bodecka-Zych, M., Gehrer, N., Krejtz, K., **Lawrence, C.**, Schönenberg, M., & Jusyte, A. (2022). Gender differences in sensitivity to provocation and hostile attribution bias toward ambiguous facial cues in violent offenders and community-based adults. *Motivation and Emotion*, 1-10.

- My contribution consisted of the following: Comment and guidance on data analysis and earlier drafts.
- Marta Bodecka-Zych's contribution consisted of: conducting the research, preliminary data processing, assistance with the first version of the article and responding to reviews - percentage contribution: 20%



" Signature

C) I declare that in the paper:

Bodecka-Zych, M., Zajenkowska, A., & **Lawrence, C**. (2022). Dad, are they laughing at me? Fathers' vulnerable narcissism and sons' hostile attributions. *Personality and Individual Differences, 192*, 111582.

- My contribution consisted of the following: Comment and guidance on data analysis and earlier drafts.
- Marta Bodecka-Zych's contribution consisted of: conceptualizing and designing the study, conducting the research, performing statistical analyses, preparing the first version of the manuscript and responding to reviews - percentage contribution: 80%

''Signature

Warszawa, 14.11.2022 r.

Imię i Nazwisko: Adrianna Jakubowska

Afiliacja: Akademia Pedagogiki Specjalnej im. Marii Grzegorzewskiej,
Szkoła Doktorska

Oświadczenie

Oświadczam, że w artykule:

Bodecka, M., **Jakubowska, A.**, Zajenkowska, A. (2021). Warsztat psychoedukacyjny z elementami mentalizacji i jego rola w redukowaniu wrogości wśród osób osadzonych.[Psycho-educational workshop with elements of mentalization and its role in reducing hostility among inmates] *Psychologia Wychowawcza, 62*(20), 101-115.

- Mój wkład polegał na: pomocy w przygotowaniu pierwszej wersji manuskryptu.
- Wkład Marty Bodeckiej-Zych polegał na: prowadzeniu warsztatów, przygotowaniu pierwszej wersji manuskryptu i odpowiedzi na recenzje – wkład procentowy: 80%

Podpis

Adrianna Jakubowska